

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:56 ; Search time 463 seconds
(without alignments)
439.155 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKQPLPLNVASGEGQI.....IQRGJHRNNAQVENGKPK 1281

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6851	100.0	1281	22 AAE02189	Human breast cancer
2	6851	100.0	1281	22 ABR47615	Breast cancer asso
3	6851	100.0	1294	23 ABJ05569	Breast cancer-asso
4	6843	99.9	1281	21 AAB00190	Breast cancer proc
5	6832	99.7	1294	21 AAB40550	Human ORF314
6	4212	61.5	785	22 AAB92512	Human protein sequ
7	286	4.2	1213	22 AAM41128	Human polypeptide
8	271	4.0	758	22 ABB64656	Drosophila melanog
9	269	3.9	499	24 ABU11707	Human MDR polypep

10	269	3.9	519	21	AAB42333	Human ORF2097
11	267.5	3.9	1224	22	RAM39342	Human polypeptide
12	256	3.7	734	22	ABB71878	Drosophila melanog
13	256	3.7	1311	22	AAW93979	Human stomach can
14	256	3.7	1311	22	AAB95671	Human protein sequ
15	251.5	3.7	533	21	AAV87745	Murine embryonic b
16	248.5	3.6	516	19	AAW70971	Human Ikaros isofo
17	247.5	3.6	950	22	ABB71271	Drosophila melanog
18	246.5	3.6	537	23	ABB79561	Mouse Daedalus pro
19	241	3.5	537	15	AAAR46964	Peptide with Ikar
20	240	3.5	419	23	ABP69811	Human polypeptide
21	239.5	3.5	1296	23	ABG61925	Prostate cancer-as
22	238	3.5	461	17	AAAR92015	Human Ikaros prote
23	238	3.5	461	19	AAW72672	Human Ikaros. Hom
24	238	3.5	461	23	AAO18605	Human Ikaros prote
25	237.5	3.5	519	23	AAAC18606	Murine Iki protein
26	237	3.5	461	19	AAW70964	Human Ikaros isofo
27	235.5	3.4	1181	22	AAB95553	Human protein sequ
28	233.5	3.4	518	17	AAAR92017	Murine Ikaros prot
29	233.5	3.4	518	19	AAW72674	Mouse Ikaros m-k-1
30	233.5	3.4	518	19	AAW79966	Mouse Ikaros isofo
31	231	3.4	3572	23	ABG95659	Human nuclear acid
32	230.5	3.4	568	15	AAAR46965	Ikaros zinc finger
33	228	3.3	351	23	AAU96738	Human cytokine rec
34	225	3.3	334	17	AAAR92020	Ikaros protein. N
35	225	3.3	334	19	AAW72677	Ikaros isoform. U
36	225	3.3	334	19	AAW70969	Human Ikaros prote
37	224.5	3.3	432	17	AAAR92016	Murine Ikaros prot
38	224.5	3.3	432	19	AAW72673	Mouse Ikaros m-k-3
39	224.5	3.3	432	19	AAW70965	Mouse Ikaros isofo
40	223.5	3.3	1696	22	ABG10886	Novel human diagr
41	222	3.2	323	22	AAU15994	Human novel secret
42	222	3.2	323	22	AAU16142	Human novel secret
43	222	3.2	323	24	ABU55063	Human novel polype
44	222	3.2	323	24	ABU55211	Human novel polype
45	220	3.2	757	21	AAAB40777	Human ORF541

ALIGNMENTS

RESULT 1
AAEC2189
ID .AAE02189 standard; Protein; 1281 AA.
XX AAE02189;
XX
XX 06-AUG-2001 (first entry)
XX
XX Human breast cancer specific gene-3 (BCSG-3) protein.
XX
XX Human; breast cancer specific gene-3; BCSG-3; cytostatic; vaccine;
XX breast cancer; therapeutic; gene therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 426 /note= "Encoded by AA"
XX
XX Misc-difference 1100 /note= "Encoded by A"
XX
XX WC200137779-A2.
XX
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WQ-US32056.
XX
XX 23-NOV-1999; 99US-0166973.
XX
XX (DIAD-) CIADEXUS INC.
XX
XX Salceda S, Caferkey R, Recipon H, Sun Y;

XX WPI: 2001-367602/38.
 DR N-PSDB; AAD06861.
 XX
 PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
 PT imaging, preventing and treating cancers, particularly breast cancer -
 XX
 XX Claim 2; Page 61-66; 66pp; English.
 XX
 CC The invention relates human breast cancer specific genes (BCSG's) and
 CC their corresponding proteins. BCSG is useful for diagnosing, staging,
 CC monitoring, imaging, preventing and treating breast cancers. BCSG is also
 CC useful for inducing an immune response against a target cell expressing
 CC BCSG. The invention also provide methods for detecting genetic lesions or
 CC mutations in BCSG, thereby determining if a human with the genetic lesion
 CC is at risk for breast cancer or has breast cancer. BCSG antibodies
 CC labelled with paramagnetic ions or radioisotopes is useful for imaging
 CC breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is
 CC useful for treating breast cancer. BCSG is useful in the rational design
 CC of new therapeutics for imaging and treating cancers. BCSG is also used
 CC in gene therapy. The present sequence is human breast cancer specific
 CC gene-3 (BCSG-3) protein.
 XX
 SQ Sequence 1281 AA;

Query Match 100.0%; Score 6851; DB 22; Length 1281;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRKNPPLRNVASEGQILEPITGESKVGKNKEFSADQMSENTDQSDAEALNHKEH 60
 DB 1 MVRKNPPLRNVASEGQILEPITGESKVGKNKEFSADQMSENTDQSDAEALNHKEH 60

QY 61 SLHVQDPSSSKKDLKSAVLSEKAGNYESPSPKGNFPSPFPHDEVTDNRKYLAFSPAAG 120
 DB 61 SLHVQDPSSSKKDLKSAVLSEKAGNYESPSPKGNFPSPFPHDEVTDNRKYLAFSPAAG 120

QY 121 VCEPLKSPQAEADDDQDMACFSGDSLETKEDQKMSPKATEGTGAQSQANCOGLSPV 180
 DB 121 VCEPLKSPQAEADDDQDMACFSGDSLETKEDQKMSPKATEGTGAQSQANCOGLSPV 180

QY 181 SVASKNPQVPSDGGVRLNKSKTLLVNDNPDAPLSPESQDFKCNCGYGYGNDPTDJI 240
 DB 181 SVASKNPQVPSDGGVRLNKSKTLLVNDNPDAPLSPESQDFKCNCGYGYGNDPTDJI 240

QY 241 KHRPKYHLGLHNRTRQDAELDSKILALHNMVQFSKDFQKNRSVFSGLQDINSRPV 300
 DB 241 KHRPKYHLGLHNRTRQDAELDSKILALHNMVQFSKDFQKNRSVFSGLQDINSRPV 300

QY 301 LLNGTYDVQVTSGGTFIGIGRKTPTDCQNTKYPKCFKCNFTYMGNSSTELCHFLQTHN 360
 DB 301 LLNGTYDVQVTSGGTFIGIGRKTPTDCQNTKYPKCFKCNFTYMGNSSTELCHFLQTHN 360

QY 361 KIKASLPSSSEVAKPSEKSNKSI PALQSSDSGLGKMQCKITVKAGDDTPGVSYVPIKPL 420
 DB 361 KIKASLPSSSEVAKPSEKSNKSI PALQSSDSGLGKMQCKITVKAGDDTPGVSYVPIKPL 420

QY 421 DSSRQNGTEATSYVCKFKCFSCSSSSKLLBHYGKHGAVQSGGKLNPELNDKLSRGSV 480
 DB 421 DSSRQNGTEATSYVCKFKCFSCSSSSKLLBHYGKHGAVQSGGKLNPELNDKLSRGSV 480

QY 481 INQNDLAKSSEGETMTKTDKSSSGAKKDFSSKGAEDNMVTSYNCQPCDFRYSKSGPVP 540
 DB 481 INQNDLAKSSEGETMTKTDKSSSGAKKDFSSKGAEDNMVTSYNCQPCDFRYSKSGPVP 540

QY 541 IVVGPLLRHYQQLHNTHKCTIKHCPFCPRGLCSPEKHLGEITYPPFACRKNCSHCAALLL 600
 DB 541 IVVGPLLRHYQQLHNTHKCTIKHCPFCPRGLCSPEKHLGEITYPPFACRKNCSHCAALLL 600

QY 601 HLSFGAAGSRVRKHQCHQCSFTTTPDQVLLFHYESVHESQASDVQKQEAHNLQSGDQCSV 660
 DB 601 HLSFGAAGSRVRKHQCHQCSFTTTPDQVLLFHYESVHESQASDVQKQEAHNLQSGDQCSV 660

QY 661 KESKESCTKCDFITOVEEBEISRHRYRAHSCYKRCQCSFTAAADTCSLLEHNTVHCQEQD 720
 DB 661 KESKESCTKCDFITOVEEBEISRHRYRAHSCYKRCQCSFTAAADTCSLLEHNTVHCQEQD 720

QY 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKVGPEVSESVVKEKLEKDGLEKVV 780
 DB 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKVGPEVSESVVKEKLEKDGLEKVV 780

QY 781 WTSSDDLNRVTRGADILRGSPSYTOASLGLLTPVSGTQEQTKTLRDSNPVAAHLAR 840
 DB 781 WTSSDDLNRVTRGADILRGSPSYTOASLGLLTPVSGTQEQTKTLRDSNPVAAHLAR 840

QY 841 PIYGLAVETKGFLOGAPAGGKSGALPOQVPASGENSKSDSQSLRRRSGGVFCANCL 900
 DB 841 PIYGLAVETKGFLOGAPAGGKSGALPOQVPASGENSKSDSQSLRRRSGGVFCANCL 900

QY 901 TTKTSLMRKNANGGVYCNACGLYOKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQA 960
 DB 901 TTKTSLMRKNANGGVYCNACGLYOKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQA 960

QY 961 EQLNKQQRGSNEEOVNGSPLERRSEDLHTESHQREIPLPSLSKYEAQGSILTKSHSAQOPV 1020
 DB 961 EQLNKQQRGSNEEOVNGSPLERRSEDLHTESHQREIPLPSLSKYEAQGSILTKSHSAQOPV 1020

QY 1021 LVSQTLDIHKEMQPLHIQIKSPQESTGDPGNSSSVSEKSGSERGSPTEKYMRPAKHNY 1080
 DB 1021 LVSQTLDIHKEMQPLHIQIKSPQESTGDPGNSSSVSEKSGSERGSPTEKYMRPAKHNY 1080

QY 1081 SPPGSPIEKYQYPLFGPLPFVHNDPQSEADMLRFWSKYL SVPGNPHYL SHVPGLPNQCQ 1140
 DB 1081 SPPGSPIEKYQYPLFGPLPFVHNDPQSEADMLRFWSKYL SVPGNPHYL SHVPGLPNQCQ 1140

QY 1141 YVPYPTNLPHPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPNVNSGPLNVV 1200
 DB 1141 YVPYPTNLPHPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPNVNSGPLNVV 1200

QY 1201 KTEKVDSTQDELSCTKVHCGIVFDEVMYALHVSCHGDSGPFQCS:COHLCTDKYDFTT 1260
 DB 1201 KTEKVDSTQDELSCTKVHCGIVFDEVMYALHVSCHGDSGPFQCS:COHLCTDKYDFTT 1260

QY 1261 HIQRLHNNACVSKGPKKE 1281
 DB 1261 HIQRLHNNACVSKGPKKE 1281

RESULT 2
 ABR47615
 ID ABR47615 standard; Protein; 1281 AA.
 XX ABR47615;
 XX 12-JUN-2003 (first entry)
 XX Breast cancer associated protein sequence SEQ ID NO: 472.
 XX Human; breast cancer; cytostatic; gene therapy.
 XX Homo sapiens.
 XX WO2003004989-A2.
 XX 16-JAN-2003.
 XX 21-JUN-2002; 2002WO-US-9669.
 XX 21-JUN-2001; 2001US-299887P.
 XX 27-JUN-2001; 2001US-301572P.
 XX 18-JUL-2001; 2001US-308501P.
 XX 25-SEP-2001; 2001US-325002P.
 XX 05-MAR-2002; 2002US-362585P.
 XX 14-MAY-2002; 2002US-380391P.
 XX (MILL-) MILLENIUM PHARM INC.

AC AAB40550;
XX 09-FEB-2001 (first entry)
XX Human ORFX ORF314 polypeptide sequence SEQ ID NO:628.
DE
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO20058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-027728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX MP1; 2000-602362/57.
XX N-PSDB; AAC74759.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 750-753; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 316. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy.
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 1294 AA;

Query Match 99.7%; Score 6832; DB 21; Length 1294;
Best Local Similarity 99.8%; Pred. No. 0;

	Natches	1279;	Conservative	0;	Mis-matches	2;	Indels	0;	Gaps	0;
QY	1	MVRKKNPP	L	RNVASEGEGQ	L	PIG	TESKVS	GKNKEFS	ADOMSENTDQ	SDAAE
DB	14	MVRKKNPP	L	RNVASEGEGQ	L	PIG	TESKVS	GKNKEFS	ADOMSENTDQ	SDAAE
QY	61	SLHVQDPSSSK	DL	KA	VLSEKAGFN	YES	SPSGKGN	PPSP	PHDEVTD	RNNLAFS
DB	74	SLHVQDPSSSK	DL	KA	VLSEKAGFN	YES	SPSGKGN	PPSP	PHDEVTD	RNNLAFS
QY	121	VCEPLKSP	Q	AEADDP	Q	DM	ACT	PSG	DSLET	KEQKMS
DB	134	VCEPLKSP	Q	AEADDP	Q	DM	ACT	PSG	DSLET	KEQKMS
QY	181	SVASKNP	Q	VP	SDGVR	L	NK	SKTDL	L	VNDNDP
DB	194	SVASKNP	Q	VP	SDGVR	L	NK	SKTDL	L	VNDNDP
QY	241	KHFRXYH	L	GLHNR	TRQDAE	L	DSK	L	ALHNMV	Q
DB	254	KHFRXYH	L	GLHNR	TRQDAE	L	DSK	L	ALHNMV	Q
QY	301	LJNGTYD	V	Q	VTSGT	R	IG	K	TPC	COGNTK
DB	314	LJNGTYD	V	Q	VTSGT	R	IG	K	TPC	COGNTK
QY	361	KIKASLP	S	SE	VAKP	SE	K	N	K	S
DB	374	KIKASLP	S	SE	VAKP	SE	K	N	K	S
QY	421	DSSRQNG	T	E	A	T	S	Y	Y	K
DB	434	DSSRQNG	T	E	A	T	S	Y	Y	K
QY	481	INONDLA	K	S	S	E	G	E	T	M
DB	494	INONDLA	K	S	S	E	G	E	T	M
QY	541	IVGQPL	L	R	H	Y	Q	L	N	I
DB	554	IVGQPL	L	R	H	Y	Q	L	N	I
QY	601	HLSPGAA	G	S	S	R	V	K	H	Q
DB	614	HLSPGAA	G	S	S	R	V	K	H	Q
QY	661	KESKEHS	C	T	K	D	P	I	T	O
DB	674	KESKEHS	C	T	K	D	P	I	T	O
QY	721	TTANGEED	G	H	A	I	S	T	T	K
DB	734	TTANGEED	G	H	A	I	S	T	T	K
QY	781	WTESSSD	L	R	N	V	T	R	G	A
DB	794	WTESSSD	L	R	N	V	T	R	G	A
QY	841	PYGLAVET	K	G	F	L	O	G	A	P
DB	854	PYGLAVET	K	G	F	L	O	G	A	P
QY	901	TTKTS	L	R	K	N	G	A	G	Y
DB	914	TTKTS	L	R	K	N	G	A	G	Y
QY	961	EQLNKQ	R	G	S	N	E	E	C	V
DB	974	EQLNKQ	R	G	S	N	E	E	C	V
QY	1021	LVSOTL	D	I	H	K	N	O	P	L
DB	1034	LVSOTL	D	I	H	K	N	O	P	L

QY 1081 SPGPSPIEKYOYPLGFLPFVHNDFOSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQ 1140
 Db 1094 SPGPSPIEKYOYPLGFLPFVHNDFOSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQ 1153
 QY 1141 YVPYPTNLPHFSAVGSDNDIPDLDAIKHSRPGPTANGASKETKAPPNVKNEGPLNVV 1200
 Db 1154 YVPYPTNLPHFSAVGSDNDIPDLDAIKHSRPGPTANGASKETKAPPNVKNEGPLNVV 1213
 QY 1201 KTEKYDRSTQDELSTKCHVGIGVFLDEVYVALHMSCHGDSGPFQCSICQHLCTDKYDFTT 1260
 Db 1214 KTEKYDRSTQDELSTKCHVGIGVFLDEVYVALHMSCHGDSGPFQCSICQHLCTDKYDFTT 1273
 QY 1261 HIQGLHRNNAQVEKNGPKKE 1281
 Db 1274 HIQGLHRNNAQVEKNGPKKE 1294

RESULT 6
 AAB92512
 ID AAB92512 standard; Protein; 785 AA.
 XX
 AC AAB92512;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 XX Human protein sequence SEQ ID NO:10640.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0181767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto G;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WP; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10640; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 785 AA;
 QY
 Query Match 62.5%; Score 4212; DB 22; Length 785;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 782; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 495 MTKTDKSSGAKKOPFSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLH 554
 Db 1 MTKTDKSSGAKKOPFSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLH 63
 QY 555 NIHKCTIKHCFPCPRGLCSPEKHLGEITYPPACRKSNCSCALLLLHLSPGAAGSSRVKH 614
 Db 6 NIHKCTIKHCFPCPRGLCSPEKHLGEITYPPACRKSNCSCALLLLHLSPGAAGSSRVKH 120
 QY 615 QCHQCSFTTDPVDVLLFHYESVHESQASDVQAEANHLQSGDQSQSVKESKHSCTKCDFI 674
 Db 121 QCHQCSFTTDPVDVLLFHYESVHESQASDVQAEANHLQSGDQSQSVKESKHSCTKCDFI 180
 QY 675 TOVEEBEISRHRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQEQDITTTANGEEDCHALS 734
 Db 181 TOVEEBEISRHRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQEQDITTTANGEEDCHALS 240
 QY 735 TIKESPKIDFRVYNLLTPDSKQGEVPSVESVVRKLEEKQGLKEKVMWTESSDDLRNVTW 794
 Db 241 TIKESPKIDFRVYNLLTPDSKQGEVPSVESVVRKLEEKQGLKEKVMWTESSDDLRNVTW 800
 QY 795 RGADILRGSPSYTQASLGLLTPVSGTQEQTKTLDSPNVVEAAHLARPIYGLAVETKGFQ 854
 Db 301 RGADILRGSPSYTQASLGLLTPVSGTQEQTKTLDSPNVVEAAHLARPIYGLAVETKGFQ 360
 QY 855 GAPAGGEKSGALPQQYPASGENKSKDESQSLRRRSGSVFQANCLTTTSLWRKNANGG 914
 Db 361 GAPAGGEKSGALPQQYPASGENKSKDESQSLRRRSGSVFQANCLTTTSLWRKNANGG 420
 QY 915 YVCNACGLYOKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQAECLNKOQSGSNEEQ 974
 Db 421 YVCNACGLYOKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQAECLNKOQSGSNEEQ 480
 QY 975 VAGSPLEERSSEDLTESHOREIPLPSLSKYEAQGSLSKSHSAQQPVLSQTLDIHKRYQP 1034
 Db 481 VAGSPLEERSSEDLTESHOREIPLPSLSKYEAQGSLSKSHSAQQPVLSQTLDIHKRYQP 540
 QY 1035 LHQIKSPQESTGDPONSSSVSEKSGSSERGSPIEKYMPAKHPNYPSPGSPIEKYQYPL 1094
 Db 541 LHQIKSPQESTGDPONSSSVSEKSGSSERGSPIEKYMPAKHPNYPSPGSPIEKYQYPL 600
 QY 1095 FGLPFVHNDFOSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQNYVPYPTNLPHFS 1154
 Db 601 FGLPFVHNDFOSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQNYVPYPTNLPHFS 660
 QY 1155 AVGSDNDIPDLDAIKHSRPGPTANGASKETKAPPNVKNEGPLNVVVKTEKVDSTQDELS 1214
 Db 661 AVGSDNDIPDLDAIKHSRPGPTANGASKETKAPPNVKNEGPLNVVVKTEKVDSTQDELS 720
 QY 1215 TKVCHGCVFLDEVYVALHMSCHGDSGPFQCSICQHLCTDKYDFTTHIQGLHRNNAQVE 1274
 Db 721 TKVCHGCVFLDEVYVALHMSCHGDSGPFQCSICQHLCTDKYDFTTHIQGLHRNNAQVE 780
 QY 1275 KXGK 1278
 Db 781 KXWK 784
 RESULT 7
 AAB41128


```

Db 936 COMH---PDL-----RNSLTGFRVCVMQVTSTJELKIHGTFRMOKL----- 975
Qy 1048 DPGNSSVSEKGGSSERSGPIEKYMRPAKHFNYPSPGSPIEKYQVPLFGLPFVHNDPQSE 1107
Db 976 --AGSSAAS-----SPNGGGLQK-----LYKCALCLKEFRSK 1005
Qy 1108 ADWLRFWSKYKLSVPGNPHYL-----SHVPGI--PNPQNYVPYPTFNLPPHFSA 1155
Db 1006 QDLV-----KLVNGLFPGLCAGCMARSANGVGLAPPADR--PCAGLRCPGCSVK 1057
Qy 1156 VGSNDIPDLAIKHSRPGPTANGASKETKAP--PNVKNQEPNVNVKTEKVDRTQ----- 1210
Db 1058 FESAEDLESHMQVDHRLJTPETSGPRGTQTSPVRKTYQCIKQMTFENEREIQHVA 1117
Qy 1211 -----DELSTKCVHCGIVFLDEVMYALHM-----SCHGDSGPPQCSIC-----OHL 1251
Db 1118 NHMIEGINHECKLQNMQFDPKALLCHLIEHSFEGMGGTFCPCVCTVFQANKLOH: 1177
Qy 1252 -----CTDKYDFTTHIQRLHNNNAQ 1272
Db 1178 FAVHGOEDKIYDCSQCPQKFFFTQELQNTHTMSHAQ 1213

RESULT 8
AB064656
ID AB064656 standard; Protein; 758 AA.
AC AB064656;
XX
XX
Dc 26-MAR-2002 (first entry)
XX
XX
Dc Drosophila melanogaster polypeptide SEQ ID NO 20760.
XX
Kw Drosophila; developmental biology; cell signalling; insecticide;
Kw pharmaceutical.
XX
Cs Drosophila melanogaster.
XX
Pn WC20017:042-A2.
XX
XX
Pd 27-SEP-2001.
XX
Pf 23-MAR-2001; 2001WO-US09231.
XX
Pr 23-MAR-2000; 2000US-191637P.
XX
Pr 11-JUL-2000; 2000US-0614150.
XX
Pa (PSE) PE CORP NY.
XX
Pi Venter JC, Adams M, Li PWD, Myers EW;
XX
Dr N-PSDB; ABL08759.
XX
XX
Pt New isolated nucleic acid detection reagent for detecting 1000 or more
Pt genes from Drosophila and for elucidating cell signalling and cell-cell
Pt interactions.
XX
Ps Disclosure; SEQ ID NO 20760; 21pp + Sequence Listing; English.
XX
Cc The invention relates to an isolated nucleic acid detection reagent
Cc capable of detecting 1000 or more genes from Drosophila. The invention is
Cc useful in developmental biology and in elucidating cell signalling and
Cc cell-cell interactions in higher eukaryotes for the development of
Cc insecticides, therapeutics and pharmaceutical drugs. The invention
Cc discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
Cc sequences (ABL01840-ABL16175) and the encoded proteins
Cc (ABB57737-ABB72072).
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic format directly from WIPO
Cc at ftp.wipo.int/pub/published_pct_sequences.
XX

```

```

SQ Sequence 756 AA;
Query Match 4.0%; Score 271; DB 22; Length 758;
Best Local Similarity 19.7%; Pred. No. 2; I.e.12;
Matches 156; Conservative 97; Mismatches 301; Indels 248; Gaps 29;

Qy 551 QQLNHHTKTIKHCPFCRGCLCSPEKH-----LGIITYPFACRKNCSHCALJLLHLSPG 605
Db 115 QQHHEHHHLMGGFNPLTPEGLNPNVQHFFYGGNLRSPQPTPTASTIAPVAV----- 166
Qy 606 AAGSRVKVQCHQCCSPTTPDQVLLPFHYESVHESASQVUKQEAHLQGSDDGQGVK---S 662
Db 167 AGSS-----EKQALTPPMQV--PPASPAKSSQSN-EPEKEHDQXNSSEDKYMAE 218
Qy 663 SKEH-----SCTKCDFTT 675
Db 219 SEDDDTNIKMPIYNHSGMKNYKCKTCGVVAITKVDFWAHTRTHMKPKILCCPKCFVT 278
Qy 676 QVEEEISRYR--RAHSCYKQCQCSFTAAQTOSLLEHNT-----VHCQEQIIT-- 723
Db 279 EFKHLEYHIRKHQKQKFCQDKCSYTCVNMJNSHRKSHSVYQYRCACQDIATKYCH 338
Qy 724 ---ANGEEDGHAISTIKER-----PKIDFRVTNLT--PDSKMGEPVSESVVRKLEKRD 774
Db 339 SFKJLHURKYGHKPGMVLDEGTFNPSLVIDVYTRGPKSKNGGPIASG----- 387
Qy 775 GLKEKYWTSSSDDLRNVTWRGADILRGSPSYTQASLGLLTPVSGTQEQTKTRDSPNVE 834
Db 388 -----GSGSGS-----RKSNAVAAPQOQOQSPA--QFVA-TSQLSAAALCGFLVQ 430
Qy 835 ---AAHLARPYYGLAVETKGFLOGAPAGKEKGALPQYYPASGENKSKCESLRRRG 891
Db 431 GNSAPPAAGSVLPJ-----PASPAKSVASVECTPSLPSP----- 464
Qy 892 SGVFCANCJTXTLSWRKANGVYCNACGLYQKLSHTPRNLNLIKQNGEQIIRRTRK 951
Db 465 -----ANLPLPLASLLQQRNMAFT-----PWNLN-----LQMLAAQOCAAVAL-----A 504
Qy 952 RLNFEALQAEQLNKKQSGNEFVNGSPLEKRSSECHLTESHQREIPLPALSKEYAQSIT 1011
Db 505 QLSPRMEEQLQQNQCCSDNEEEQDDEYERKSVDSAMDJ-----SGTFV 550
Qy 1012 KHSACQPVLSOTLDIHKRMOPJHIQIKSPOESTGDPONSSSVSEGGKSSERGSPIEKY 1071
Db 551 KEDEQQ-----QPCQLAVNKLKVEEAT--PLMSSSNASRRKG---RVLKLDL 595
Qy 1072 MRPAKHPNYSP-----PGSPIEKYQVPLFG---LPEVNDPQSEADWLFEWYKLSVSG 1123
Db 596 LQLRSEAVTSPQIKVPTPTTASPIAGRPXPEEHCSTSSAD----- 641
Qy 1124 NPHVLSHFGLRNFCCQYVVEYPTFNLPPHFSAVGSDNDIFDLAIKHSRPGPTANGASKE 1183
Db 642 ESMETAHVPCANTSASSTA-----SSGNSSNASSNS 673
Qy 1184 KTKAPNVNKGEPNVVVKTEKVDRTQDELSKCVHCGIVFLDEVYALHMSCHGDSGPF 1243
Db 674 KGNSSSNSSNGTTSVAAPPSGTAAAGATYECKYCDIFFKDAVLYTHMGVHSCDDVF 733
Qy 1244 QCSICQHLCTDKYDFTTHIQRLH 1267
Db 734 KCMWCEKCGGPGVGLFVHMARNAH 757

RESULT 9
ABU11707
Xc ABU11707 standard; Protein; 499 AA.
XX
XX ABU11707;
XX
Dc 13-FEB-2003 (first entry)
XX
Dc Human MDET polypeptide SEQ ID 654.
XX

```

KW MDDT; human; disease detection and treatment molecule polypeptide;
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
 KW gene therapy; protein replacement therapy; cell proliferative disorder;
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
 KW psoriasis; hepatitis.

OS Homo sapiens.
 XX WO2000279449-A2.
 XX 10-OCT-2002.
 XX 27-MAR-2002; 2002WO-US09944.
 XX 28-MAR-2001; 2001US-279613P.
 XX 29-MAR-2001; 2001US-280067P.
 XX 16-MAY-2001; 2001US-280068P.
 XX 17-MAY-2001; 2001US-291280P.
 XX 17-MAY-2001; 2001US-291829P.
 XX 17-MAY-2001; 2001US-291849P.
 XX 19-JUN-2001; 2001US-299428P.
 XX 20-JUN-2001; 2001US-299776P.
 XX 20-JUN-2001; 2001US-300001P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;
 XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 XX Petalta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX WPI: 2003-058431/05.
 XX N-PSDB; ABX34697.

XX New purified disease detection and treatment molecule proteins and
 XX polynucleotides, useful for diagnosing, treating or preventing cancers
 XX (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
 XX or hepatitis -

XX Claim 27; SEQ ID NO 654; 339pp + Sequence Listing; English.

XX This invention describes a novel disease detection and treatment molecule
 XX polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
 XX osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
 XX antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
 XX and the polypeptides of the invention can be used for gene therapy,
 XX protein replacement therapy and are useful for treating a variety of
 XX diseases or conditions. These polypeptides or polynucleotides are
 XX particularly useful for diagnosing, treating or preventing cell
 XX proliferative disorders (e.g. cancers including adenocarcinoma,
 XX leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
 XX disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
 XX syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
 XX hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
 XX by ABU11450-ABU11845, described in the disclosure of the invention.
 XX NOTE: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 499 AA;

XX Query Match 3.9%; Score 269; DB 24; Length 499;
 XX Best Local Similarity 25.9%; Pred. No. 1.5e-12;
 XX Matches 139; Conservative 67; Mismatches 192; Indels 38; Gaps 29;

QY 814 LTPVSGTQEQTKLRDPSNVEAHMLRPIYGLAVETKGFLOGAPA---GGKSGALPQQY 870
 DB 17 LSTTSGGQSSK---SDRVASN-----VKVETQSDENGACEMNGEEDRLNL 65

QY 871 PASGENKS---KDSQSLLRRRRSGV-----FCA-NCLTTK-TSLMRXANG 913
 DB 66 DASSEKMGSHRDQSSAL-----SGVGIRLPNGKCLKDCGIIICGPNVLMVHKRSH 120
 QY 914 G---YVNCACGL-----YOKLMS-PPPIWIKQNGEQIIRRTKRLNPEALCA 960
 DB 121 GERPFCCQCGASFTQKGNLRLHKLHSGEXP---FKHLCNYACRRR-----DALTG 170
 QY 961 EQLNCKQSGSHNEQVNGSPJERRS--EDHLTSHQ--REIPLPS-----LSKYEAGGS--- 1009
 DB 171 HLRTHSVCKPHKCGYCGRSYKQRSJEEHKERCHNYLESNGJPGTLYPVKEETHSEYA 230
 QY 1010 --LTXSHSAQCPVLVNSQTLQ-HKEMQPLHIQIKPQESTGDPGNSSSVSEKGSSEKSP 1067
 DB 231 EDLCKIGSERSLVLDRLASNVAKKSSM-----PKFLGDKGLSDTPYDSSASVEKENE 284
 QY 1068 -EK-----YMRPAKHFNYSFPG-----SPIEKYQYPLP-GLPFV 1100
 DB 285 YMKSHVMOQA-NNA-NYLGASLSRLPVG---TPPGSGSEWVPI-SPMYQJHKPLAEG-FRS 341
 QY 1101 HNDQFSEA-EWLRKFSKYKLSVPGKPHYLSHVPGLP-NPCQNVVPYPTFNLPHFSAVGS 1158
 DB 342 YHSAQCSAVENLLLSKAKL-VP-----SREASPSNCCSTCTESNNEEQSGLIYL 394
 QY 1159 ENDFELDLAKHSRSGPTANGAS-KEKTKA-----PPNVKNEGPLNVVYKTEKVRSTQDE 1212
 DB 395 -NNEI-----APHAR-----NGLSLKEHRAVDLL-RAASENSQDALRVVST-----SGEQM 439
 QY 1213 LSTKVCVCGIYFLEVMYALHWSCHGSDGSPFOCSIQHJCTCKYDFTTHICSGYHR 1268
 DB 440 KYVKCEHRCRVLFLOHVMYTIHYGCHGSRDPFECNMCGVHSQDRVEFSHITRGEHR 495

Human ORFX ORF2097 polypeptide sequence SEQ ID NO:4194.

Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; neurotic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacteria; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; ashma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WC200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WC-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0545763.

XX (CURA-) CURAGEN CORP.


```

QY 48 QSDAAELNHEKEHSLH-----VDDP-----SSSKKDLKS--AVLSE--KAGFNYESP 91
Db 13 QODFESLADLTDRRAHRCPGDGDQDDPQLSWVASPSKSDVASPTQMGDCDLGLGEEEG 72
QY 92 SKGDNFSPFHDEVTDRNMLAFSPAAGVCEPLKSPQRAEADDPQDMACTPGSDSLETK 151
Db 73 GTGLUPFC-----QFCDSFRLSY-----LKRHEQIHS-DKLPFKCTYCSRLFKKH 118
QY 152 EDQKMSPKATEETQCAQSGOANCOGLSPVSVASKNPQVPSDGGVRLNKSXTDLLVNDNDP 211
Db 119 RSRDRHKL--HTGDKKYHCECAAFS-----RSDHLXHLKXTHSSSK 150
QY 212 PAPLSPBLOPKFCNICGYGYNDPTDLIKHR-----KYHLGLHNRT----- 254
Db 161 P-----FKCTVCKRGF--SSTSLQSHMQAHKKKKEHLAKSEKAKKDDFMCDYCE 209
QY 255 ---RODAELOSILALHNYQFHSKDFOKV-----NSLAVLEIHLKTIHADKPGQSHTQCIC 293
Db 210 DTFSQTELEKHLVLRH--POLSEKADLQCHCEPEVFVDENTLLAHFHQAHANQKHCPM 257
QY 284 ---RSVFSGLVDINSR-----PVL-----LNGTYD--VQVTSGGT----- 315
Db 268 CPEQFSSVEGYVCHLDSHRQDSSNHSVSPDVLGVSAMSSSATPDSSASVERGSTPDS 327
QY 316 ---FISGRKTPDCQGNK-YFRCKFCNFTYMGNSSTELSOHFLQTHPNKKA----- 354
Db 328 LKPLRGQKMRDQGGWTKVYSPCYCKRDF-NSLAVLEIHLKTIHADKPGQSHTQCIC 396
QY 365 --SLPS---SEVAKPSEKSNKSI PALQSSD-----SGDLGKWKDKITVK-A 405
Db 387 LDSMPTLYNLNHEHVRKLUK--NHAYPMQFCNLSAFHCNCPENFAFADINSLOH:RVSHC 444
QY 406 GDDTPVGVSPKPLDSSRQNGTEATSYMKCFCSFSCSSSLKLEHYCKQHGAVCSG 455
Db 445 GPNA-----NPSDGNNAFFCQCSMGFLTSS--JTEHIQCAHCSVGA 486
QY 466 GLN-----PELNDKLSRGSVI-----NQNDL-----AKSEGETMT 496
Db 487 KLESFVVQPTOSFMEVYSCPYCTNSPIFGSILKLTKHKENHKNPLAHSKSKAEQSPV 546
QY 497 KTKSSSGAKKKDFSSKGAEDNMVTS--YNCFCDFRYSKSHGPDVIVVPLLRHYQCLH 554
Db 547 SSDVEVSPKQRULS-----ASANSISNGEYPCNQCDLKFSPFES-----FQTHLKLH 594
QY 555 NIHKCTTIKHCFPCPRGLCSPEKHGE-TYPPACRKSX--CSHC-----ALLL 600
Db 595 LELLRLKQACPOCKEDFDSQBSLLQHLTVMYMTTS:HYVCESCDKQFSSVDDLLQKHLLOX 654
QY 601 HLSPGAAGSRVHKHQHCQCSFTTPDPDVLJFHYESVHESQAS-----DYKQEAHL 651
Db 655 H-----TFVLVHCTLCOEVPDSKVSIOVHLAVRHSNEKMYRCTACNWDFRKEA--- 703
QY 652 QGSDGQSVKES-----KEHSCTKCDFI TOVEBEISRHYRRAHSCKYKQCQCSFTAADTQ 705
Db 704 ---DLQVHVHSHLGNPAKAHKCIFCGEETSTVELOCHITTHSKVYKCRFCSKAFHAI 760
QY 706 SLLEHFNTHVHQEQDITITANGEDG-HAISTIKKEPKIDFRVYNLLTPDSKNGEYSESV 764
Db 761 LLEKHLREKHC-VFDAATENGANGVPMATKKAEP-ADLOGLLLKNPEA----PNSH-- 812
QY 765 VKREKLEKGLKEKVTWESSDDL-RNVTWRGADIIRGSPSYTQASLGLLTPVSGTCEQ 823
Db 813 -----EASEDDVDASBPYMGCDIC--GAAYTMEVJ-----LQ 842
QY 824 TKTLRSPNVEAHLARPIYGLAVETKGFLOGAGGKSGALPQOYPASGENSKSDESO 893
Db 843 NHRURD-----HNRIP-----GEDDGRSKKAEFFYKSHKCNVCSR 877
QY 884 SLLRRRSGGVFCANCILTTKTSIWRKNANGVYVNCAGLYQKLCSTPRPLNI:KQNGEQ 943
Db 878 TF-----FSENGLEHLOTHRGPAK-HYMCPIG-----ERFSLTLTJEH----- 917

```

RESULT 12

AB971878
ID ABB71879 standard; Protein: 734 AA.

XX ABB71879;

XX AC

XX DC 26-MAR-2002 (first entry);
XX DE Drosophila melanogaster polypeptide SEQ ID NC 42426.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WC260171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US03231.

XX PR 22-MAR-2000; 2000US-191537P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX (PEXE) PE CORP NY.

XX XX Venter JC, Adams M, Li PWD, Myers EW;

XX XX WPI; 2001-656960/75.

XX DR N-PSDB; ABL15981.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions .

XX PS Disclosure; SEQ ID NC 42426; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL31840-ABL316175) and the encoded proteins

```
CC (AB57737-ABB2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from W-PO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 734 AA;
Query Match 3.7%; Score 256; DB 22; Length 734;
Best Local Similarity 19.4%; Pred. No. 3.3e-11;
Matches 175; Conservative 118; Mismatches 303; Indels 304; Gaps 37;
QY 444 ESSSSLKLEHYGKHGAVQSGGLNPELNDKLSRGSVINQNDLAKSSGEMTKTDKSSS 503
DB 7 KGNHLKASELYGTCH--ERTGG-----KFGQRIYVQWE-----QDTTQQQQQQQ 50
QY 504 GAKKDFSKGAEDNMVTSYQCQCD-FRYSKSHGPQVIVVGPV-----LRHYQLHNH 557
DB 51 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 110
QY 558 KCTIKH--CPCPRGLCSPERHGLGITYPFACRK-----SYCSHCA 596
DB 111 AATSVGVVVTQAGAGVSEPEQY--VVVPRNQRILLT--AGTLENEAREGEPPS--NAGNA- 167
QY 597 LLLLHLSPGAGSSRVKHQ--CHQCSFTTPDVVLLFHVSVHESQAS-DVQKQANHLQG 653
DB 168 -----SSGSASDSHIEVQSAHQSPGATHVQMAPRNAEVTQGEATGESQQQQQQQQ 221
QY 654 SDGQQ-----SVKSEKHSCTKCDFTQVEEIEISRHYRRASHCYKC 694
DB 222 HQHQQQQQCPTPNGASYGETTIVISSEAEALQH-----HHQQQQHQQQCHQCHQ 275
QY 695 RCCSFTADTQSL--EHFNTVHCOEDI--TTANGEDGHAISTIKEEPIKIDFV----- 746
DB 276 AAASAAAQTVHIAATSGHGTFRVTFEDVRFTAGPETS--ASNMYVDVVPVDSVHANE 323
QY 747 -----XNLLTPDS-----KMGEPV-----SESVKREKLEEK 773
DB 334 SKTYADLGNAYFPFSSFSFSSNSYAATLQGGNTIYVPTGQFLAKSESGJCTGLRQ 393
QY 774 DGLK-----EKWTESSDDLRYNTWR-----GADILGSPSYQAS 810
DB 394 TGPATFQITISFEGNGIEPLMASPAPPEYQVCFNPFHQVIDEYSGNMTSHKWPASS 453
QY 811 LCLLTPVSGTQECTYLRDSPNVEAHLARPIYGLAVET-----KG----FLOQAFAG 860
DB 454 IQQY-----DGSLVTASTSSPNHELKNCCHGPFRLKGSSEYFCNCP--- 496
QY 861 EKSGLPQQYPASGENKSKDESQSLLRRRRSGVFCANCLTKTSLWRKNANGGVVCNAC 920
DB 497 ----AFMRMAPRITQRAKPKAAAPNNR--NGVTCANQINSTLWRNNEGNPVCNAC 551
QY 921 GUYQKLHSTPRPLNTIKQNGEQIITRRTRKRLNPEALQAEQLNQQKQSGNEEQVNSGP 980
DB 552 GUYLKLHNMNRLPSKKEG-----IQKRRKPKNN-----GGAPM 596
QY 981 ERSSEDLHTESHQREIPLPSLSKYBAQGSLLTKSHSAQQPVLVSQTLDIHKRQMLHIQIK 1040
DB 587 HR-----APLPSMS-----QGVNLMANSLYPSQVPSM-----ENSLN 622
QY 1041 SPQES-----*GDPGNSSSVSEGGKSGRSGPIEKYMRPAKHPNYSPPGSPIERYK 1091
DB 622 SQONSSPELHDSMTTQAGGQVWVSI SLNATAPPTPDGLNMSARHHVTVGSHSPYSQOS 681
QY 1092 YPLFLGLPFVHDFQSEADWLRFWSKYKLSVPGNPHYLGHVPLGNPQNYVPYPTFNLP 1151
DB 682 TP-----SQOS-----PHLPQ----- 692
QY 1152 HFSAVGSDNDIPLDLAIKHSRPGTANGASKETKAPNNVKNEGPLNVVTKTEKVDRSQD 1211
DB 693 -----TVPINQI--VQVPVTIESRSNSTELTPSV-----ITRTGLPERSNN 734
```

RESULT 13

```
AA93979
ID AA93979 standard; Protein; 1311 AA.
XX
AC AA93979;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human stomach cancer expressed polypeptide SEQ ID NO 27.
XX
KW Human; stomach cancer; marker; screening; micro-metastasis;
peritoneal dissemination.
XX
OS Homo sapiens.
XX
PN WO2000:09317-A1.
XX
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WC-CP05063.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 18-OCT-1999; 99US-0159590.
PR 11-JAN-2000; 2000JP-0118776.
PR 17-FEB-2000; 2000US-0183322.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELIX) HELIX RES INST.
XX
PI Cta T. Isozai T. Nishi-kawa T. Hayaishi K. Saito K. Yamamoto J.
PI Ishii S. Sugiyama T. Wakamatsu A. Nagai K. Otsuki T. Aburatani H.
PI Kodama T. Midorikawa Y.
XX
XX WP1: 2001-570287/64.
XX
DR N-PSDB; AA193855.
XX
PT New Stomach cancer-associated genes, useful as markers in blood tests
for screening for the early stages of the disease.
XX
XX Claim 1; Page 91-93; 242pp; Japanese.
XX
XX The invention relates to stomach cancer-expressed genes
(CC (AA193842-AA193979)) and the encoded proteins (AA193967-AA194039). The
genes can be used as markers in blood tests for screening for the early
stages of the disease. The proteins and peptides can be used as targets
for screening for compounds to treat the disease. They can also be used
for predicting micro-metastases. The gene can predict peritoneal
dissemination.
XX
SQ Sequence 1311 AA;
Query Match 3.7%; Score 256; DB 22; Length 1311;
Best Local Similarity 19.1%; Pred. No. 8.9e-11;
Matches 293; Conservative 178; Mismatches 491; Indels 570; Gaps 80;
QY 19 QLEPTG--TESKVSNNKEFADQMSNTDOSDAELNKEEHLHVQCP-----SSS 70
DB 54 QVFESLCTIEHKI-----NQCLTDGVQVED-----DPTCSWPASSP 91
QY 71 SKKDLKSAVLSKAGNYESPSSKGNFPSPFPHD-EVTDRLNMLAFSPAAAGGVCEPLKSPQ 129
DB 92 SKKQDTSPHSGECDFGEE---EGG--PCLPVPQPCCKSFSLSY-----LKHE 137
QY 130 RAEDADPQDVACTPSSGDSLETKEQKMSPKATETGTQAGSQGANCQG-----LSPV 180
DB 138 QSHS-SKLFPKCYCSRLFHKHESDRHKL-HTGDKKYHCSECDAAFRSDHLKHLK 194
QY 181 SVASKNP---QVPSDGGVRLNKSITDLLVNC-NPQAPLSPQLQDF-----KCNICGYG 230
DB 195 THSNKPYKCAICRRGFLSSSSLHGHHQVHERNKGSGSGRVEDWKMMDTKKCSQCEEG 254
QY 231 YVGNQDPTDLIKHFKYHJGLHNPTRQCA-----ELDSKILALHNMVQFSKSKQFQKN 283
```

```

255 F--DPPEDLOKHIAECPSPEDRAALQCVYCHELFVEETSLXNHEQVHSGE-KNNS 311
QY
284 RSVFSGVYQDI-----NSRPPVLLN-GTYDVQVTSSTGTFIGT----- 319
Db
312 CSICSESFHTVEELYSHMDSHQCPESCNHNSPSLVTGVYTSVSTTPDSNLVDSSTVY 321
QY
320 -----GRK-----TPDCQNTK-----YFRCKFCNFTYMGNSSTLEOHFLOTH- 358
Db
372 EAAPPIPKSRGRKRAAQOTPDMTGSSKQAKVTVSYCYCN-KQJFSSSLAVLQHLKTMHL 430
QY
359 --PNKIRASLPSSSEVAKPSEKNSKSPAL-QSSDSG-----DL 394
Db
431 DKPEQAHICQVCLEVL-PSLYNLNEHLKQVHEAQDPGLIVSAMPAYVQCNPFCSEVNDL 489
QY
395 GKWDQKITVKGADTTP-----VGYSVPKIPLDSSRON----- 426
Db
490 NTLQEHIRCSHGFPANPAKOSNAFFPCPHYMGFT-----JSSLEHIRQVHCDLSSRF 544
QY
427 -----GT-----EATSYNWKFCSPFCSSSSSLUKLEHYKQKHGAVCSGGLNPELNDKJLSPG 478
Db
545 GSPVLGTPKEPVVSVYSCYCTNSPIFNSVLKLNKHIXENHNKIPL-ALNVIHNGKKSRA 603
QY
479 SVINQNDLAKSSEGETWTKDKSSSGAKKXDFSSKGAEDNMVTSYNCQFCDFRY 533
Db
504 -----LSPSPVAIBOTSLKMQVAGGAPAP-----TGEYICNQCAKYTSLDSF 649
QY
534 -----KSHGPDVI-----VWGP-----LRLHYQQLNHIHKCTIKHCPFCRGLCSF-- 574
Db
650 QTHLKTHTLDTVPLKTPCQCNKEPPNQESLLKHV-TIHFMTITSYICESCDQFTSVDC 708
QY
575 -EKHLGEI-IYFPACRKSNSHC-----ALLHLSPGNAGSSRVKHQCHQCSFTTPD 625
Db
709 LQKHLDMHTFVF-----FRCTLCQEVFDSKVSIGLHLAVKHSNEKV-YRCTSCNWDFRN 763
QY
626 VDVLFFHYESVHESQASDVKEANHLQSDQOQSVKESKESCTKC--DFITOVEEB--I 681
Db
764 ETDLQLH-----VKH--NHLE-----NQKVKHCLIFGCSFGTEVELOCHI 802
QY
682 SRHYRAHSCVKQCQSFSTAADTOSLLEHFNTHVCQEQDITKANGEDGHAISTIKEEPK 741
Db
803 TTHSKK-----YNCKFCSKAFHAILLEKHLREKHC-VFEKTPCNGINGASEQVQKEEVE 857
QY
742 IDFRVYNLLTPDSKMGEPVSVVREKLEKGLKEKWTESSSDLRLNVTWVGADILR 801
Db
858 LQ-----TLNTN-----SOSHNHSDGSEEDVDTSE-----PMYGCQIC- 891
QY
802 GSPSYTOASGLLTPVSGTQTKLRDSPNVEAAHLARPIYGLAVETKGLQAPAGGE 861
Db
892 -GAAYTMETL-----LQNHQLRD-----HNIRP----- 913
QY
862 KSGALPOOYPASGENKSKDESQSLRRRRRSGVFCANCLATKTSMLRKNANGVYVNCAG 921
Db
914 -----GESAIYVKKAELEIK-----GNYKNVQS 936
QY
922 LYKXLTSTPAPLNIKONGEQIIRRTKRLNP-----EALQAEQLN 964
Db
937 -----RTFFSENG-----LREHMOTHLGPKVHYMCPICGERPFSLLTJTEHKVTH 981
QY
965 KQQRGSNEOVNGSPLERRRSHEDHJTESHOREIPLPS:SKVEAQGLTKSHSAQCPVVSQ 1024
Db
982 SKSLDTGNCRIKMPLO--SSEEFLEHCQM--PDL-----RNSLTGFRVCVCMQTVTS 1031
QY
1025 TLDTHKMQPLHIQIKFSQESTGDPGNSSSVSEKGSSEKSPLEKYRYFAKHNKSPFG 1084
Db
1032 TLED-KINGTTHMQ-----NTNGSAVQ-----TTGRGHVQX----- 1063
QY
1085 SPIEKYQVPLGLFVINDPOSEADWLRFNFKYKLSVPGPHFY-SHVPGLPNCQNY--V 1142
Db
1064 -----LYKASCULKEFRSKODLV-----KLIDNGLPY-----GLCAGCVNLSKS 1102
QY
1143 PYPTFNLP-----HFSAV-----GSDNIPJDLAKHS 1171

```

```

Db 1103 ASFGNVPPETNRPGLQGNENLSAIEGKGVGLTRCSCSNVKPESSELCNH:QT:HR 1162
QY 1172 RQCFANGASKSKTKAPPNKNKNEGFNVVXTEKVDPRSTODELST-KCVHCGIVF:DE-- 1227
Db 1163 ELVPSDNSSTCLXTPQSP-----MPRISPSQSDCKKTYQC:KQCMVFNEM:DI 1210
QY 1228 -WYVALHX-----SCH-----GDSGFQCSIC----- 1248
Db 1211 QVHEVANH:DEGLNHECKLCSOTFSPAKLOCHLIEHSEFGMGTFKCPVCFTVFQANK 1270
QY 1249 --QHL-----CTKYDFTTHIQ 1263
Db 1271 LQGHFSAHQSDKIYDCTCQCPQFFFTQLQ 1302

RESULT 14
AAB9557:
ID AAB9557: standard; Protein: 1311 AA.
XX
AC AAB9557:
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18457.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP107461.7-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0302253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-2181757.
PR 09-JUN-2000; 2000JP-2241595.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 8; SEQ ID :8457; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB9557 to AAB13628 and

```


XX Claim 1a; Page 22-23; 29pp; Japanese.

XX This invention describes a novel isolated Eos protein and its encoding nucleic acid. The Eos gene can be used for the development of new drugs for various cerebral nervous diseases. Independent claims are also included for the following: (1) a cloned cDNA carrying the above Eos gene and exhibiting an activity of inhibiting formation of abnormality in nervous system; (2) an Eos gene expression product having the above amino acid sequence; (3) a cloned cDNA expressing Eos protein, its fragment, its derivative and its homolog; (4) a human homolog of Eos gene; (5) an expression product of the human homolog of Eos gene; (6) inhibiting formation of abnormality in nervous system including a step of introducing Eos gene in a cell by using the above cloned cDNA; (7) a drug composition, that is a gene therapy agent, containing the above Eos gene as the active component. This represents the murine embryonic brain Eos protein described in the method of the invention.

XX Sequence 533 AA;

Query Match 3.7%; Score 251.5; DB 21; Length 533;
Best Local Similarity 18.8%; Pred. No. 4.4e-11;
Matches 154; Conservative 86; Mismatches 244; Indels 337; Gaps 32;

QY 484 NDLAKSSEGETM---TKTDKSSGAKKKDFSSKGAEDNMVTSYNCQFDPFRYSKSHGPD- 539
DB 12 SSLEKDSLGAPGVSVSTPNSQHSFSLSANSIXVEM-----YSDSESRLLGPD 64
QY 540 -----VIVVGPLLRLHYQQLNIHKTIKHCPCPRGLC---SPEKHLGEITYPFACR 568
DB 65 RLKDDESVIVEDSLSE-----PLGYCGSGGPEPH----- 94
QY 589 KSNCSHCALLLLHLSFGAGSSRVKHCQCSFTTPDQVLLPHYBS-----VHESQ-A 641
DB 95 -----SPGIRLPNGKLKCDVCGMVCIGPNVLMVHKRSHTGERPFHCNQQG 140
QY 642 SDVKQKAN---HLQSDGQGSVKESKEHSCTKDFITQVEEISRHRYRAHSC----- 691
DB 141 ASPTQKGNLLRHKLKSGEKF--CPFCNYACRRRDALTGHL-RTHSVSSPTVGK 193
QY 692 -YKROCSTAAQTQS:LEHFNVTVCQEQDITTANGEEDGHAISTIKKEPKIDFRVYN:LL 750
DB 194 PYKNCYGRSVKQGSTLEEHEKCHNYLQSLT-----DAQALT----- 232
QY 751 TPDSKMGEPVSESVVKREKLEKDKLKEKVMTESSDDLRNVTWKGADILRGS---PSYT 807
DB 233 -----GQF-----GDEIRLEWVPDSMLHPSTERPTF- 260
QY 808 QASLGILLTPVSGTQCTKTLRDSPNVEAAHLAPRYGLAVETKGFLOGAPGGEKSGALP 867
DB 261 DRLANSLT-----KRRRSTP 275
QY 868 QQYPASGENSKDBESQSLRRRRSGGVFCANCLTTKTLWRKNANGGVYCNAGLYQKLH 927
DB 276 QKP--VGEKQWRFSLSL-----PYDVNASGGY----- 301
QY 928 STPRPLNIKQNGEQIIRRRTRKRLNPEALQAEQLNKQORGSENEOVNGSP:ERRSECH 987
DB 302 -----EKDVELVAHHGLEPGF-----GGSLAFVGTET 328
QY 988 LTESHQREIPLPSLUSKYAQSGLTKSHSAQQPVLVSQTLDIHKRMQPLHIQIKSPOESTG 1047
DB 329 L-----RPLRLPP-----TNCISELTPVISS---VYTMQPIPSRLSLP----- 364
QY 1048 DPGNSSSVSGKSGSSRGSPTEKYMRPAKHPNYSPPGSP:EKYQYPLFG:PFVHNDFOSE 1107
DB 365 ---GSRAGEGPELDGGGL-----LYRARGSLTDPGASP-----SNGCQDS 404
QY 1108 ADM:RFWSKYKLSVPGNPHVLPNPNQNVVYPFTNLPHPFSAVGSDNDIPLDLA 1167
DB 405 TD---TESNHEDRIGG-----VGIPSSGSPQFPPTIVGGRHSPAYAKEDPKQGS-G 452
QY 1168 IKHSRPGPTANGASKEXTKAPPNVNREGPLNVVKTETKVDRSTQDELSTKCVHCGIVFLDE 1227

DB 453 LLRTGPGP-----SKEVLRV-----VGSSEFVKAFKCEHCISFLDH 490
QY 1228 VVVALHKSCHESGPFQCSICQHLCTKYDFTTHIQRLHR 1268
DB 491 VVFTIRKQCHSFRDPPFECNICGYSQRYEFSSHIVGGRHX 531
Search completed: October 29, 2003, 1:38:47
Job time : 465 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:57 ; Search time 26 Seconds
(without alignments)
2084.623 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851
Sequence: 1 MVRKNPNPLRVASGEGQI.....IQRLHRNNAQVKNQKPKRE 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCUS_COXB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248.5	3.6	516	3	US-08-711-417C-202
2	238	3.5	461	3	US-08-711-417C-196
3	238	3.5	461	4	US-08-733-622C-25
4	233.5	3.4	518	3	US-08-711-417C-196
5	233.5	3.4	518	4	US-08-733-622C-27
6	230.5	3.4	568	5	PCT-US93-08743-5
7	225	3.3	334	3	US-08-711-417C-201
8	225	3.3	334	4	US-08-733-622C-23
9	224.5	3.3	432	3	US-08-711-417C-197
10	224.5	3.3	432	4	US-08-733-622C-26
11	210.5	3.1	376	3	US-08-711-417C-200
12	210.5	3.1	376	4	US-08-733-622C-29
13	209.5	3.1	232	4	US-08-733-622C-31
14	208	3.0	236	3	US-08-711-417C-175
15	206.5	3.0	390	3	US-08-711-417C-199
16	206.5	3.0	390	4	US-08-733-622C-28
17	205	3.0	431	3	US-08-711-417C-195
18	205	3.0	431	4	US-08-733-622C-24
19	200	2.9	507	4	US-08-733-622C-2
20	200	2.9	976	3	US-08-894-997-50
21	195.5	2.9	236	3	US-08-711-417C-174
22	193.5	2.8	853	3	US-09-254-325-2
23	190.5	2.8	238	3	US-08-711-417C-177
24	189.5	2.8	470	3	US-08-465-590-153
25	189.5	2.8	470	2	US-08-283-300A-13
26	188.5	2.8	470	4	US-08-733-622C-22
27	188.5	2.8	470	5	PCT-US95-09345-13

28 198 2.7 449 2 US-08-927-394-2 Sequence 2, Appl.
29 187.5 2.7 470 3 US-08-711-417C-153 Sequence 153, App
30 187.5 2.7 1706 2 US-08-459-568-2 Sequence 2, Appl
31 187.5 2.7 1706 2 US-08-393-411-2 Sequence 2, Appl
32 187.5 2.7 1706 3 US-08-516-859A-2 Sequence 2, Appl
33 187.5 2.7 1706 4 US-09-586-472-2 Sequence 2, Appl
34 187.5 2.7 1706 4 US-09-528-706-2 Sequence 2, Appl
35 183.5 2.7 240 4 US-08-733-622C-30 Sequence 30, Appl
36 182.5 2.7 866 3 US-09-079-415-6 Sequence 6, Appl
37 182.5 2.7 866 3 US-08-750-458A-2 Sequence 2, Appl
38 180.5 2.6 233 4 US-08-733-622C-36 Sequence 36, Appl
39 177.5 2.6 727 2 US-08-475-844-9 Sequence 9, Appl
40 177.5 2.6 727 5 PCT-US95-08429-9 Sequence 9, Appl
41 177 2.6 728 2 US-08-475-844-5 Sequence 5, Appl
42 177 2.6 728 5 PCT-US95-08429-5 Sequence 5, Appl
43 174 2.5 1494 3 US-08-755-587-186 Sequence 186, App
44 171.5 2.5 238 3 US-08-711-417C-176 Sequence 176, App
45 164.5 2.4 676 2 US-08-398-590A-40 Sequence 40, Appl

ALIGNMENTS

RESULT 1

US-08-711-417C-202
; Sequence 202, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: PastSeq for Windows Version 2.0b
; CURRENT APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 09/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 202:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 202:
US-08-711-417C-202
Query Match 3.6% Score 248.5; DB 3; Length 516;
Best Local Similarity 25.7%; Pred. No. 5.3e-13;

LENGTH: 461
TYPE: PRT
ORGANISM: Homo sapiens
US-08-733-622C-25

Query Match
Best Local Similarity 3.5%; Score 238; DB 4; Length 461;
Matches 128; Conservative 58; Mismatches 194; Indels 116; Gaps 25;

QY 847 VETGKFLGAGAA---GGKSGALPOQYPASGENKS---KDSQSLLRRRRSGV-----894
DB 4 VETOSDEENGRACEMNGEACEDLMLDASGKNGSHRDQSSAL-----SGVGIRLP 58
QY 895 -----PCA-NCLTTKTLRWKQANGG---YVGNACGL-----YQKLHSTPRPL 933
DB 59 NGKLNKDCICGIIICGPNVLMVHKSHGTGERPQCNGGASFTQGNLRLHKLHSGEKFP 118
QY 934 -----NIIKQNGEQIIRRRTRKRLNPEALQAEQKQNGSNEEQVNGSPLEERSDHL 988
DB 119 KCHLCNVACRRRDALTGHLRTHSVGKPKHCGYGRSYKQRTSLEEH-----KERCHNYL 172
QY 989 TESHOREIPLSLSKYEAGS-----LYKSHSAQOPVLVSOTLDIHKRMQPLHIQKSPQ 1043
DB 173 -ESMGLPGLTLPVPIKEETKTHSEMAEDLCKIGSERSLVLDRLASNVAKGKSSV-----PQ 225
QY 1044 ESTGDPGNS-----SSVSEGGSSSERGSPIEKYMRPAKHPNY-----SPDGS-- 1055
DB 226 KFLGDKGLSDTYPYSATYKENEKMMKSHVMQAINNA--INYLGAESLRPLVQTPPGGSE 283
QY 1086 --PIEKYQYPLF---GJFFVHNDPQSEA-DMLRFWSKYK-SVPONPHYLSHVPGLP-NP 1137
DB 284 VVPVISPMYQLHRRSEGTFRSHSAQDSAVEYLLLSKAKL-VP-----SEREASPSNS 336
QY 1138 CONYYVPTPLPHFSAVSGNDIP-----LDLAIKHSRPGPTANGASKEKTKAPPNVK 1192
DB 337 CQSDTDSNNEBORSGIYLTNHTARRAQRVSLKEEH-RAYDLLRAASENSQDA-----390
QY 1193 NEGPLNVVYKTEKVBSTODELSTKCVHCGIYFLDEVMYALHMSCHGSGGPFCCSICQHL 1252
DB 39: ----LRVVST-----SGEOMKYKCEHCRVLFDHVMYTIHMGCHGFRDPFECNMGVHS 44:
QY 1253 TDKYDFTTHIORGLHR 1268
DB 442 QDRVEFFSHITGEHR 457

RESULT 4
US-08-711-417C-198
Sequence 198, Application US/08711417C
Patent No. 6228611
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia A.
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
NUMBER OF SEQUENCES: 202
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711,417C
FILING DATE: 05-Sep-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,212
FILING DATE: 02-MAY-1994
APPLICATION NUMBER: 08/121,438

FILING DATE: 14-SEP-1993
APPLICATION NUMBER: 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis P.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10287/007901
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 220154
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-08-711-417C-198

Query Match
Best Local Similarity 3.4%; Score 233.5; DB 3; Length 518;
Matches 126; Conservative 70; Mismatches 192; Indels 157; Gaps 25;

QY 814 LTPVSGTQCTKTLRP-SPNVEAAHLARPIYGLAVETKCFLOGAPA---GGKSGALPQ 869
DB 37 LST-SQAQNSKSDRMAASNVK-----VETQSCDENGRACVNGSECAEDLRM 84
QY 870 YPASGENKS---KDSQSLLRRRRSGVFCANCLTTKTLRWKNGVGVVGNACGL-----922
DB 85 LDASGKNGSHRDQSSAL-----SGVSGI-----RLPNGKLNKDCICGVCIG 128
QY 923 -----YQKLHSTPR-----LNIIKQNGEQ-----IRRRTRKRLNP 955
DB 129 PNVLVHKSHGTGERPQCNGGASFTQGNLRLHKLHSGEKPKKCHLCNVACRRR---185
QY 956 EALCAQLNKQKQSGNEEQVNGSPLEERS--EDHLTESHQ--REPLPSL-----BK 1003
DB 186 DALTGHRLRTHSVGKPKHCGYGRSYKQRTSLEEHKERCHNYLESMLGPGVCFVKEETH 245
QY 1004 YFAGSLTSHSAQOPVLVSOTLDIHKRYQPLHIQKSPCESTGDPGNSSSVSEGGSGSE 1063
DB 246 NEMAEELCKIGERSVLDLASNVAKPKSSM-----PQKFGKCLSDMPYDSANYEK 299
QY 1064 RGSPIEKYMRPAKHP-NY-----SPDGS-----PIEKYQY-----PLFGLPFFVH 1101
DB 300 EDNMTSHVMEQAINNAIYLGAEISLRPLVQTPPGSSEVVPV-SSMYQLHKEKPSDGPGRN 359
QY 1102 NDFQSEADKLRFWSKYKLVPGNPHYLSHVPGLP-NPCQNYVP-----YPTF 1147
DB 360 HSAQDAVDMLLLSKAK-SVS-----SEREASPSNSCQSDTDSNNEBORSGIYLTN 412
QY 1148 NLPPHPSAVGSDNDIPDLAIKHSRPGPTANGASKEKTKAPPNVKNGEPLNVVYKTEKVR 1207
DB 413 HINPH-----ARNGLALKEQRAYEVLRAASENSQDAFRVVS 449
QY 1208 STQDELST-KCVHCGIYFLDEVMYALHMSCHGDSG---PFQCSIQHLCTDKYDTTHIC 1263
DB 450 TSQEQLKVKYKCEHCRVFLDHVMYTIHMGCHGCHGFRDPFECNMGVHSQDRYEFSSHT 509
QY 1264 RGLHR 1268
DB 510 RGEHR 514

RESULT 5
US-08-733-622C-27
Sequence 27, Application US/08733622C
Patent No. 6528634
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia
APPLICANT: Norgar, Bruce A.

☆

```
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-Sep-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 201:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 201:
US-08-711-417C-201

      Query Match      3.3%; Score 225; DB 3; Length 334;
      Best Local Similarity 24.2%; Pred. No. 3e-11;
      Matches 94; Conservative 53; Mismatches 145; Indels 96; Gaps 17;

QY 915 YVCNACGL-----YQKLHSTPRPLNIKQ-----NGEQIARRTRK 951
DB 5 FQNCQGASFTQGNLLRHILKHSGEKPKCHLCNYACRRDALTGHLRTHSVIKEETK- 63
QY 952 RLNPEALQAEQLNK--QQRGSNEEOVNGSPLEERS---EDHLTESHQSEIPLPLSLSKYEA 1006
DB 64 ----HSEMAEDLCIKGSESLVLDRLASNVAKRKSMPQKFLGDKGLSDTPYDS-ATYEK 118
QY 1007 QGSLTKSHSAQGPVLVVSQTLDIHKRMQPLHIQIKSPQESTGDPGNSSSVSEKGSSEGRS 1066
DB 119 ENEMKSHVMDQAINNAINYLGAESLRPL---VQTP-----PG-----GS 155
QY 1067 PIEKYMRPAKHPNYSPPGSPKIEKYQYPLFGLPFVHNDQFSEADMLRFMSKYKLSVPGNPH 1126
DB 156 EVVPVISWYQLHRRSEGTGRSN-----HSAQDSAVEYLLLSKAKJ-VP-----199
QY 1127 YLSHVVGLP-NPCQNVYPYPTFNLPHPFSAVGSDNDIP-----LDLAKHSRPGPTANGA 1180
DB 200 --SEREASPSNCSQDSTJTESNNEORSGLIYLTNHIARRAQRVSLEKEH-RAYDYLRAA 256
QY 1181 SKEKTAPPNVKNVEGPLNVVKTETKVDRTQDELSTKCVHCGIVFLDEVYALHMSCHGDS 1240
DB 257 SENSQDA-----LRVVS-----SGEQMKVYKCEHCRVFLDHWYTYTHYGGCHGR 302
QY 241 GPFQCSICHLCTDKYDFTTHIQRGJHR 1268
DB 303 DPPECNMGYHSQDRYEFSSHITRGEHR 330

      RESULT 8
      US-08-733-622C-23
```

```
; Sequence 23, Application US/08733622C
; Patent No. 6528634
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: A10LOS GENE
; FILE REFERENCE: 10287-030001
; CURRENT APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 60/017,646
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 60/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-733-622C-23

      Query Match      3.3%; Score 225; DB 4; Length 334;
      Best Local Similarity 24.2%; Pred. No. 3e-11;
      Matches 94; Conservative 53; Mismatches 145; Indels 96; Gaps 17;

QY 915 YVCNACGL-----YQKLHSTPRPLNIKQ-----NGEQIARRTRK 951
DB 5 FQNCQGASFTQGNLLRHILKHSGEKPKCHLCNYACRRDALTGHLRTHSVIKEETK- 63
QY 952 RLNPEALQAEQLNK--QQRGSNEEOVNGSPLEERS---EDHLTESHQSEIPLPLSLSKYEA 1006
DB 64 ----HSEMAEDLCIKGSESLVLDRLASNVAKRKSMPQKFLGDKGLSDTPYDS-ATYEK 118
QY 1007 QGSLTKSHSAQGPVLVVSQTLDIHKRMQPLHIQIKSPQESTGDPGNSSSVSEKGSSEGRS 1066
DB 119 ENEMKSHVMDQAINNAINYLGAESLRPL---VQTP-----PG-----GS 155
QY 1067 PIEKYMRPAKHPNYSPPGSPKIEKYQYPLFGLPFVHNDQFSEADMLRFMSKYKLSVPGNPH 1126
DB 156 EVVPVISWYQLHRRSEGTGRSN-----HSAQDSAVEYLLLSKAKJ-VP-----199
QY 1127 YLSHVVGLP-NPCQNVYPYPTFNLPHPFSAVGSDNDIP-----LDLAKHSRPGPTANGA 1180
DB 200 --SEREASPSNCSQDSTJTESNNEORSGLIYLTNHIARRAQRVSLEKEH-RAYDYLRAA 256
QY 1181 SKEKTAPPNVKNVEGPLNVVKTETKVDRTQDELSTKCVHCGIVFLDEVYALHMSCHGDS 1240
DB 257 SENSQDA-----LRVVS-----SGEQMKVYKCEHCRVFLDHWYTYTHYGGCHGR 302
QY 1241 GPFQCSICHLCTDKYDFTTHIQRGJHR 1268
DB 303 DPPECNMGYHSQDRYEFSSHITRGEHR 330

      RESULT 9
      US-08-711-417C-197
      Sequence 197, Application US/08711417C
      Patent No. 6228611
      GENERAL INFORMATION:
      APPLICANT: Georgopoulos, Katia A.
      TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
      NUMBER OF SEQUENCES: 202
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: Fish & Richardson P.C.
      STREET: 225 Franklin Street
      CITY: Boston
      STATE: MA
      COUNTRY: USA
      ZIP: 02110-2804
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      COMPUTER: IBM Compatible
      OPERATING SYSTEM: Windows 95
```

SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711,417C
FILING DATE: 05-Sep-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,212
FILING DATE: 02-MAY-1994
APPLICATION NUMBER: 08/121,438
FILING DATE: 14-SEP-1993
APPLICATION NUMBER: 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis P.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10287/00700:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 197:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 197:
US-08-711-417C-197

Query Match 3.3%; Score 224.5; DB 3; Length 432;
Best Local Similarity 22.3%; Pred. No. 5.1e-11;
Matches 114; Conservative 66; Mismatches 156; Indels 175; Gaps 22;
QY 814 LTPVSGTQEQTKLRD-SPNVEAAHARPIYGLAVETKGFQGAPA---GGEKSGALPQQ 869
DB 37 LSTTSGAQQNSKDRGVASNYK-----VETQSDDENGRACEMNGEECAEDLRM 84
QY 870 YPASEGNKSKS---KDESQSLRRRRSGVFCANCLTTKTLARKVANGGYVCNACGL--- 922
DB 85 LDASGEKNGSHRQGGSSAL-----SGVGGI-----RLPNGKJKCDICIGVICG 129
QY 923 -----YQKLHSTPRP-----LNLIKQNNGEQ-----IIRRTKRLNP 955
DB 129 PNVLVWVKHRSHTGERPFCQNSGASFTQKGNLRLHKLHSCEKPFKCHLCNACRR--- 185
QY 956 EALQAEQLNKQCRGNEECVNGSPLERRSEDLHTEHQCRIPPLSLSKYEAQGSJLKSHS 1315
DB 186 -----DALTGHLRTHSGCKCLSDMPVDS-ANYEKEDNMT-SHV 221
QY 1016 AQPVLVSQTLDIHKRQVPLHIQIKSPQESTGDPGNSSSVSEGGKSGSERGSPIEKYMRPA 1075
DB 222 MDQAINNAIYLGAEISLRPL---VQTP-----PGSEVVPVVISXYQLH 262
QY 1076 KHPNYSPPGSP:EKYQYPLFGLPFVHNDFOSEADWLRFSKYKLSVPGNPHYLSHVPGLP 1135
DB 263 KPPSDGPPRS-----NHSQAQDAVGNLLLSKAK-SVS-----SEREASP 300
QY 1136 -NPCQNYVP-----YPTFNLPHPFSVAGSNDIPLDLAIKHSRPGPTANGAS 1181
DB 301 SNSQDSTDTESNAEEQRSGLIYLTNHNPH-----ARNGLA 337
QY 1182 KEKTAAPPNVKNGEPLNVVTKEDVDRSTQDEJST-KCVHCGIVFLDEVWVAJHNSCHGDS 1240
DB 338 LKEEQRAYEVLRAAENSQDAFRVSVTSGEOLKYVKECHRCVLPFDHVMYTHMGCHGCH 397
QY 1241 G---PFOCSICQHLCTDKYDFTTHIQRLGHR 1268
DB 398 GFRDPFECNMGYHSQDRYEFSSHITRGEHR 429

RESULT 10
US-08-733-622C-26
; Sequence 26, Application US/08733622C

Patent No. 6528634
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia A
APPLICANT: Morgan, Bruce A
TITLE OF INVENTION: A-GLOS GENE
FILE REFERENCE: 10287-033001
CURRENT APPLICATION NUMBER: US/08/733,622C
CURRENT FILING DATE: 1996-10-17
PRIOR APPLICATION NUMBER: US 60/017,646
PRIOR FILING DATE: 1996-05-14
PRIOR APPLICATION NUMBER: US 60/005,529
PRIOR FILING DATE: 1995-10-18
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 432
TYPE: PRT
ORGANISM: Mus musculus
US-08-733-622C-26
Query Match 3.3%; Score 224.5; DB 4; Length 432;
Best Local Similarity 22.3%; Pred. No. 5.1e-11;
Matches 114; Conservative 66; Mismatches 156; Indels 175; Gaps 22;
QY 814 LTPVSGTQEQTKLRD-SPNVEAAHARPIYGLAVETKGFQGAPA---GGEKSGALPQQ 869
DB 37 LSTTSGAQQNSKDRGVASNYK-----VETQSDDENGRACEMNGEECAEDLRM 84
QY 870 YPASEGNKSKS---KDESQSLRRRRSGVFCANCLTTKTLARKVANGGYVCNACGL--- 922
DB 85 LDASGEKNGSHRQGGSSAL-----SGVGGI-----RLPNGKJKCDICIGVICG 129
QY 923 -----YQKLHSTPRP-----LNLIKQNNGEQ-----IIRRTKRLNP 955
DB 129 PNVLVWVKHRSHTGERPFCQNSGASFTQKGNLRLHKLHSCEKPFKCHLCNACRR--- 185
QY 956 EALQAEQLNKQCRGNEECVNGSPLERRSEDLHTEHQCRIPPLSLSKYEAQGSJLKSHS 1315
DB 186 -----DALTGHLRTHSGCKCLSDMPVDS-ANYEKEDNMT-SHV 221
QY 1016 AQPVLVSQTLDIHKRQVPLHIQIKSPQESTGDPGNSSSVSEGGKSGSERGSPIEKYMRPA 1075
DB 222 MDQAINNAIYLGAEISLRPL---VQTP-----PGSEVVPVVISXYQLH 262
QY 1076 KHPNYSPPGSP:EKYQYPLFGLPFVHNDFOSEADWLRFSKYKLSVPGNPHYLSHVPGLP 1135
DB 263 KPPSDGPPRS-----NHSQAQDAVGNLLLSKAK-SVS-----SEREASP 300
QY 1136 -NPCQNYVP-----YPTFNLPHPFSVAGSNDIPLDLAIKHSRPGPTANGAS 1181
DB 301 SNSQDSTDTESNAEEQRSGLIYLTNHNPH-----ARNGLA 337
QY 1182 KEKTAAPPNVKNGEPLNVVTKEDVDRSTQDEJST-KCVHCGIVFLDEVWVAJHNSCHGDS 1240
DB 338 LKEEQRAYEVLRAAENSQDAFRVSVTSGEOLKYVKECHRCVLPFDHVMYTHMGCHGCH 397
QY 1241 G---PFOCSICQHLCTDKYDFTTHIQRLGHR 1268
DB 398 GFRDPFECNMGYHSQDRYEFSSHITRGEHR 429
RESULT 1:
US-08-711-417C-200
; Sequence 200, Application US/08711417C
Patent No. 6228631
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia A
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
NUMBER OF SEQUENCES: 202
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
City: Boston

STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711,417C
FILING DATE: 05-SEP-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,212
FILING DATE: 02-MAY-1994
APPLICATION NUMBER: 08/121,438
FILING DATE: 14-SEP-1993
APPLICATION NUMBER: 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis P.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10287/007001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-08-711-417C-200

Query Match 3.1%; Score 210.5; DB 3; Length 376;
Best Local Similarity 22.3%; Pred. No. 7e-10;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

QY	814	LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGLQGAPA---GGEKSGALPQQ	869
DB	37	LSTTSGAQQNSKDRGYASNVK-----VETQDEENGRACENMGCECABDLRM	84
QY	870	YPASGENKS---KDESQLRRRRGSGVFCANCLT*KTSLMRKNANGGVVCHACGLYQKL	926
DB	85	LDASGEKMGSHRDQGSAL-----SGVGGI-----RLPNGKLKCDICGV---	125
QY	927	HSTPRPLNIKQNGEQIPRRTRKRLNPEALQAEOLNKQQRGSNEEQVNGSPLEERSD	986
DB	126	-----CIGNVLM-----VHKRS---	138
QY	987	HLTESHQREIPLPSLKYEAQGSLLTKSHSAQQPVLVSQT*LDHKRMQPLHIQIKSPOEST	1046
DB	139	HTGDKCLSDMPYDS-ANYEKEDWMT-SHYMDQAINNAIYNLGAESLRPL---VQTE---	189
QY	1047	GDPGNSSVSEGKSGSERGSPIEKYMRPAKHPNYSPGSPIEKYQYPLFGLFVHNDPQS	1106
DB	190	-----PCGSEVVPVISSMYQLHKPPSDGPPRS-----NHSAGQ	222
QY	1107	EADWLRFWSKYKLSVPCGNPHYLSHVPGLP-NPCQNYVP-----YPTFNLPHP	1152
DB	223	AVDNLILLSKAK-SVS-----SERASPSNSQDSTD*TESNAEQRSGLIY*TNHINPH	275
QY	1153	FSAVGSDND*PLDLA*KHSPGPTANGASKETKAPPNVKNEGFLNVVTKYKVRSTQDE	1212
DB	276	-----ARNGLALKEQRAVEVLRAASENSQDAFRVVS*SGEQ	312
QY	1213	LST-KVHCQGVFLDEWYALHMSCHDGSQ---PFCGSCIQHLCTCKYDFTTHIQRLGHR	1268
DB	313	LKVYKCHCRVFLDHYMT*IHMGCHGCHGRDPPFCNMGYHSQDRYEFSSHITRGEHR	372

RESULT 13

US-08-733-622C-31
Sequence 31, Application US/08/733622C
Patent No. 6528634
GENERAL INFORMATION:
APPLICANT: Geotopoulos, Katia
APPLICANT: Morgan, Bruce A.
TITLE OF INVENTION: AIOLOS GENE
FILE REFERENCE: 10287-030001
CURRENT APPLICATION NUMBER: US/08/733,622C
CURRENT FILING DATE: 1996-10-17
PRIOR APPLICATION NUMBER: US 60/017,646
PRIOR FILING DATE: 1996-05-14

RESULT 12
US-08-733-622C-29
Sequence 29, Application US/08/733622C
Patent No. 6528634
GENERAL INFORMATION:
APPLICANT: Geotopoulos, Katia
APPLICANT: Morgan, Bruce A.
TITLE OF INVENTION: AIOLOS GENE
FILE REFERENCE: 10287-030001
CURRENT APPLICATION NUMBER: US/08/733,622C
CURRENT FILING DATE: 1996-10-17
PRIOR APPLICATION NUMBER: US 60/017,646
PRIOR FILING DATE: 1996-05-14
PRIOR APPLICATION NUMBER: US 60/005,529
PRIOR FILING DATE: 1995-10-18
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 376
TYPE: PRT
ORGANISM: Mus musculus
US-08-733-622C-29

Query Match 3.1%; Score 210.5; DB 4; Length 376;

Best Local Similarity 22.3%; Pred. No. 7e-10;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

QY	814	LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGLQGAPA---GGEKSGALPQQ	869
DB	37	LSTTSGAQQNSKDRGYASNVK-----VETQDEENGRACENMGCECABDLRM	84
QY	870	YPASGENKS---KDESQLRRRRGSGVFCANCLT*KTSLMRKNANGGVVCHACGLYQKL	926
DB	85	LDASGEKMGSHRDQGSAL-----SGVGGI-----RLPNGKLKCDICGV---	125
QY	927	HSTPRPLNIKQNGEQIPRRTRKRLNPEALQAEOLNKQQRGSNEEQVNGSPLEERSD	986
DB	126	-----CIGNVLM-----VHKRS---	138
QY	987	HLTESHQREIPLPSLKYEAQGSLLTKSHSAQQPVLVSQT*LDHKRMQPLHIQIKSPOEST	1046
DB	139	HTGDKCLSDMPYDS-ANYEKEDWMT-SHYMDQAINNAIYNLGAESLRPL---VQTE---	189
QY	1047	GDPGNSSVSEGKSGSERGSPIEKYMRPAKHPNYSPGSPIEKYQYPLFGLFVHNDPQS	1106
DB	190	-----PCGSEVVPVISSMYQLHKPPSDGPPRS-----NHSAGQ	222
QY	1107	EADWLRFWSKYKLSVPCGNPHYLSHVPGLP-NPCQNYVP-----YPTFNLPHP	1152
DB	223	AVDNLILLSKAK-SVS-----SERASPSNSQDSTD*TESNAEQRSGLIY*TNHINPH	275
QY	1153	FSAVGSDND*PLDLA*KHSPGPTANGASKETKAPPNVKNEGFLNVVTKYKVRSTQDE	1212
DB	276	-----ARNGLALKEQRAVEVLRAASENSQDAFRVVS*SGEQ	312
QY	1213	LST-KVHCQGVFLDEWYALHMSCHDGSQ---PFCGSCIQHLCTCKYDFTTHIQRLGHR	1268
DB	313	LKVYKCHCRVFLDHYMT*IHMGCHGCHGRDPPFCNMGYHSQDRYEFSSHITRGEHR	372

; PRIOR APPLICATION NUMBER: US 60/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Gallus gallus
US-08-733-622C-31

Query Match 3.1%; Score 209.5; DB 4; Length 232;
Best Local Similarity 24.8%; Pred. No. 3.7e-10;
Matches 71; Conservative 43; Mismatches 99; Indels 73; Gaps 9;

QY 995 EIPPLSLKSYAQSGLTKSHSAQPVLSQTLDIHKMOPHIOIKSPQESTGPGKSSS 1054
DB 4 DLPYDATTNYKENENIMQTHVIDQAINNAISYLAESLRPL---VQTP----- 48
QY 1055 VSEKGSSESGSPIEKYMRPAKHPNYPSPGSPBKYQVPLFGLPFVHNDQF----- 1105
DB 49 -----PVGSEVVPV-SPMYQLHKPHGDNQTRSNHTAQD 81
QY 1106 SEADWLRFWSKYKLSVPCNPHYLSHVPLGLP-NPCNYVPVPTFNLPHFSAVGSNDIPL 1164
DB 82 SAVENLLLSLAK-SVS-----SERDASPSNSCOD-----STDTSNNEERS 122
QY 1165 DLAIKHSRPGPTA-NGASKEKTKAPPNVKNKGPLNVVKTCKVDSTQDELST-KCVHCGI 1222
DB 123 GLIVLTNIGHPHARGISVKEESRQFVLRAGTNSQDAFKVISSNGEQVYVYKCEHCV 182
QY 1223 VFLDEWYALHMSCHDGGPQCSCIQHLCTDKYDFTTHIQRG-LHR 1268
DB 193 LFLDHVMYTIHMGCHGRDPPFECNMGVYHSQDRYEFSSHITRGEHR 228

RESULT 14

US-08-711-417C-175
; Sequence 175, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-Sep-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 175:
US-08-711-417C-175

Query Match 3.0%; Score 208; DB 3; Length 236;
Best Local Similarity 26.6%; Pred. No. 5.1e-10;
Matches 72; Conservative 35; Mismatches 106; Indels 58; Gaps 9;
QY 1004 YEAGQSLTKSHSAQPVLSQTLDIHKMOPHIOIKSPQESTGPGNSSSVSEKGSSE 1063
DB 15 YEKENEMKSHVMDQAINNAISYLAESLRPL---VQTP-----PGSEVV----- 57
QY 1064 RSPIEKYMRPAKHPNYPSPGSPBKYQVPLFGLPFVHNDQFSEA-DWLRFWSKYKLSVP 1122
DB 58 -----PVISPMYQLHRRSEGTFRSNHSAQNSAVEYLLLSLAKLEKK 99
QY 1123 GNPHYLSHVPLGNPCNYPVPTFNLPHFSAVGSNDIP-----LDLAIKHSRPGPTA 1177
DB 100 KYPSERESAPS--NSCQSTDTESNNEEQRSGLYLTNHIARRAQRVSLKEEH-RAYDLL 156
QY 1178 NGASKEKTKAPPNVKNKGPLNVVKTCKVDSTQDELSTKCVHCGIVFLDEWYALHXSCH 1237
DB 157 RAASENSQDA-----LKWST-----SGEQMKVYKCEHCVFLDHWYTIHWGCH 202
QY 1238 GDSGFPQCS-COHLCTDKYDFTTHIQRG-LHR 1268
DB 203 GPRDFECNMGVYHSQDRYEFSSHITRGEHR 233

RESULT 15

US-08-711-417C-199
; Sequence 199, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-Sep-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 199:


```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 390 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: internal
;   SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-08-711-417C-199

Query Match      3.0%; Score 206.5; DB 3; Length 390;
Best Local Similarity 24.0%; Pred No. 1.7e-C9;
Matches 106; Conservative 58; Mismatches 159; Indels 119; Gaps 20;

QY 864 GALPOQYPASGENKKSQESQSLRRRRSGVFCAN----CLTTKTSLWR--KNANGG--- 914
Db 27 GDEPMFVPEDLSTTSGAQONSQSDRMGMRPFQCNQSGASFCKGNLLRHKLHSGEKPF 86
QY 915 --YVCN-ACGLYOKLHSTPRPLNIKQ--NNGEQIIRRTKRLNPEALQAEOLNK--QQ 567
Db 87 KCHLCNYACRRRDALTGHURTHSVIKEETHNE-----XAEDELCKIGAE 130
QY 968 RGSNEECVNGSPLEERS---EDHLTESHOREIPLPSLKYEAQGSUTKSHSAQAPVLVSC 1024
Db 131 RSLVLRSLASNYAKRKSMPQKFLGDKCLSDMPYDS-ANYEKEDNMT-SHYWDQALNNAI 188
QY 1025 TLDIHKMOPLHIQIKSPQESTGDPGNSSSVSEKGSERGSPIEKYMRPAKHPNYSPPG 1094
Db 189 NYLGAESLRPL---VQTP-----FGSEVVPVISMYQLHKPESDGP 229
QY 1085 SPIEKYQYPLFGLPFVHNDFOSEADWLEFWSKYKLSVPCNPHYLSHVPGLP-NPCQNVVP 1142
Db 230 S-----NHSAQDAVDNLLLSKAK-SVS-----SEREASPSNSCCDS 267
QY 1144 -----YPTFNLPPIHFSAVGSDNDPLDLAIKHSEPGPTANGASKENTKAPPN 1190
Db 268 TESNAEEQRSGLIYLTNHINPH-----ARNGLAKESQRAYE 304
QY 1:91 VKNEGPLNVVVKTEKDRSTCDELST-KCVHCGIVFLDEVNVALHMSCHCDSG---PFQCS 1246
Db 305 VLRAAENSODAFRVVSTSGEQLVKYKCEHCRVFLDHVMYTIHMSCHGCHGFRDFFECN 364
QY 1247 ICOHLCTDKYDFTTHIQRLHR 1268
Db 365 MCGYHSCDRYEFSSHITRGEHR 386
```

Search completed: October 29, 2003, 11:39:21
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:39:25 ; Search time 230 Seconds
(without alignments)
953.209 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKKNPPLRNVAESGGQT.....IQRLHNRNAQVKNKGPKR 1281

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 642050 seqs, 171146064 residues

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6851	100.0	1281	15	US-10-177-293-472 Sequence 472, App
2	246	3.6	537	14	US-10-037-667-1 Sequence 1, Appl
3	242	3.5	515	14	US-10-037-667-4 Sequence 4, Appl
4	240	3.5	419	15	US-10-037-667-321 Sequence 321, App
5	238	3.5	461	10	US-09-755-830-38 Sequence 38, Appl
6	233.5	3.4	518	10	US-09-755-830-40 Sequence 40, App
7	225	3.3	334	10	US-09-755-830-43 Sequence 43, Appl
8	224.5	3.3	432	10	US-09-755-830-39 Sequence 39, App
9	222	3.2	323	10	US-09-764-864-947 Sequence 947, App
10	222	3.2	323	10	US-09-764-864-1095 Sequence 1095, Ap
11	217.5	3.2	294	10	US-09-764-864-1515 Sequence 1515, Ap
12	215	3.1	442	10	US-09-749-728B-11 Sequence 11, Appl
13	211.5	3.1	1125	10	US-09-974-298-114 Sequence 14, App
14	210.5	3.1	376	10	US-09-755-830-42 Sequence 42, Appl
15	209	3.1	474	15	US-10-177-293-174 Sequence 174, App

16	209	3.1	2783	10	US-09-816-669A-14 Sequence 14, Appl
17	207.5	3.0	757	15	US-10-014-799A-4 Sequence 4, Appl
18	207	3.0	443	15	US-10-177-293-176 Sequence 176, App
19	206.5	3.0	390	10	US-09-755-830-41 Sequence 41, Appl
20	206	3.0	498	14	US-10-037-667-5 Sequence 5, Appl
21	205	3.0	431	10	US-09-755-830-37 Sequence 37, Appl
22	204	3.0	517	12	US-10-032-585-7874 Sequence 7874, Ap
23	201	2.9	507	14	US-10-037-667-3 Sequence 3, Appl
24	198	2.9	532	14	US-10-037-667-2 Sequence 2, Appl
25	196	2.9	481	10	US-09-764-864-973 Sequence 979, App
26	196	2.9	1154	15	US-10-232-561-10 Sequence 10, Appl
27	195	2.8	876	10	US-09-801-368-44 Sequence 44, Appl
28	194.5	2.8	135	10	US-09-764-864-983 Sequence 983, App
29	193.5	2.8	279	15	US-10-106-698-5351 Sequence 5351, Ap
30	188.5	2.8	470	10	US-09-755-830-13 Sequence 13, Appl
31	186.5	2.8	600	10	US-09-764-864-957 Sequence 957, App
32	188	2.7	1051	15	US-10-097-346-79 Sequence 79, Appl
33	187.5	2.7	1756	15	US-10-024-453-2 Sequence 2, Appl
34	187.5	2.7	1756	15	US-10-142-650-3 Sequence 3, Appl
35	187	2.7	532	10	US-09-801-368-360 Sequence 360, App
36	186.5	2.7	1258	15	US-10-198-070-101 Sequence 101, App
37	185	2.7	3067	12	US-09-949-029-18 Sequence 18, App
38	184.5	2.7	730	10	US-09-801-368-126 Sequence 126, App
39	182.5	2.7	549	10	US-09-801-368-358 Sequence 358, App
40	182.5	2.7	1506	15	US-10-232-561-4 Sequence 4, Appl
41	182	2.7	1127	15	US-10-232-561-12 Sequence 12, App
42	179.5	2.6	464	10	US-09-842-777-10 Sequence 10, App
43	179.5	2.6	1523	12	US-10-205-215-135 Sequence 135, App
44	176.5	2.6	1042	15	US-10-097-346-83 Sequence 83, Appl
45	174.5	2.5	1551	9	US-09-864-761-35904 Sequence 35904, A

ALIGNMENTS

RESULT :

US-10-177-293-472
; Sequence 472, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Heerssch, Sebastian
; APPLICANT: Xorahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
; FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/293,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 62/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 62/362,585
; PRIOR FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 472
; LENGTH: 1281
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-472

Query Match 100.0%; Score 6851; DB 15; Length 1281;
Best Local Similarity 100.0%; Pred. No. C;
Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRKNPPLRNVASEGQILEPIGTESKVGSKNKEFSADQKSENDDQSDAAELNKEEH 60
DB 1 MVRKNPPLRNVASEGQILEPIGTESKVGSKNKEFSADQKSENDDQSDAAELNKEEH 60

QY 61 SLHVDPSSSSKXDLKSAVLSEKAGFNYESPSKGNFSPFPHDEVTRNMLAFSPAAG 120
DB 61 SLHVDPSSSSKXDLKSAVLSEKAGFNYESPSKGNFSPFPHDEVTRNMLAFSPAAG 120

QY 121 VCEPLKSPQAEADDPDMACTSGSLSTKEQKSPKATETGQAQSGQANCOGLSPV 180
DB 121 VCEPLKSPQAEADDPDMACTSGSLSTKEQKSPKATETGQAQSGQANCOGLSPV 180

QY 181 SVASKNPQVPSDGGVRLNKSKTDLVNDNPDPAPLSPQLQDFKCNICGYGNGPDTLI 240
DB 181 SVASKNPQVPSDGGVRLNKSKTDLVNDNPDPAPLSPQLQDFKCNICGYGNGPDTLI 240

QY 241 KHRKXHLGLHNRTRDAELDSKILAHNMVQSHSKDFQKNRSPVSGVLQDINSRPV 300
DB 241 KHRKXHLGLHNRTRDAELDSKILAHNMVQSHSKDFQKNRSPVSGVLQDINSRPV 300

QY 301 LINGTYDVQVTSQGTFTIGIKRKYPCDCGNTKYFRCKFNFTYMGNSSTLEQHFLOTHN 360
DB 301 LINGTYDVQVTSQGTFTIGIKRKYPCDCGNTKYFRCKFNFTYMGNSSTLEQHFLOTHN 360

QY 361 KIKASLPSEVAKPSEKNKSKPALQSSDSGLKMWQKITVKGADDTFVGSVPKRP 420
DB 361 KIKASLPSEVAKPSEKNKSKPALQSSDSGLKMWQKITVKGADDTFVGSVPKRP 420

QY 421 DSSQNGTEATSYWCKFCSPCESSSLKLEHYGKHCAVQSGNLNDELNDKLSRGSV 480
DB 421 DSSQNGTEATSYWCKFCSPCESSSLKLEHYGKHCAVQSGNLNDELNDKLSRGSV 480

QY 481 INQNDLAKSSGETMTKTDKSSGAKKDFSKGAEDNMVTSYNCCFCDFRYSKSGPVP 540
DB 481 INQNDLAKSSGETMTKTDKSSGAKKDFSKGAEDNMVTSYNCCFCDFRYSKSGPVP 540

QY 541 IVGFLLRHYOQLNHNKTIKHCPGPRGLCSPEKHLGEBITYPFACRKNCSHCAJLL 600
DB 541 IVGFLLRHYOQLNHNKTIKHCPGPRGLCSPEKHLGEBITYPFACRKNCSHCAJLL 600

QY 601 HLPSCAAGSSRVKHQCHQCSFTTPDVLVLFVHESVHESQASDVQKQANHLQSDGQCV 660
DB 601 HLPSCAAGSSRVKHQCHQCSFTTPDVLVLFVHESVHESQASDVQKQANHLQSDGQCV 660

QY 661 KESKEHCTKCDFTQVBEELSHRYRAHSCYKRCQCSFTTAQTSLEHFNTHCQEQD 720
DB 661 KESKEHCTKCDFTQVBEELSHRYRAHSCYKRCQCSFTTAQTSLEHFNTHCQEQD 720

QY 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVAKRELEKXGK 780
DB 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVAKRELEKXGK 780

QY 781 WTESSDDLNRVWRGADILRGSPSYTQASLGLLTPVSGTQTKTLRDSPNVAAHLAR 840
DB 781 WTESSDDLNRVWRGADILRGSPSYTQASLGLLTPVSGTQTKTLRDSPNVAAHLAR 840

QY 841 PIYGLAVETKFLQAPAGSEKSGALPQVPPASGENKSKDESOLLRRRGSGVFCAN 900
DB 841 PIYGLAVETKFLQAPAGSEKSGALPQVPPASGENKSKDESOLLRRRGSGVFCAN 900

QY 901 TTKTSWRKXANGGYVCNACGYQKLSHTPRLNATIKONNGECIIRRTTRKRLNPEALQA 960
DB 901 TTKTSWRKXANGGYVCNACGYQKLSHTPRLNATIKONNGECIIRRTTRKRLNPEALQA 960

QY 961 EQLNKQCRSNEQVNGSPLEPRSEDHILTESHQREIPPLPSLSKYEAQSLTKSHSAQOPV 1020
DB 961 EQLNKQCRSNEQVNGSPLEPRSEDHILTESHQREIPPLPSLSKYEAQSLTKSHSAQOPV 1020

QY 1021 LVSQTLDIHKRMQPLHIQIKSPQESTGPGKSSSVSECKSGSERGSPIEKYVRAKHPNY 1080
DB 1021 LVSQTLDIHKRMQPLHIQIKSPQESTGPGKSSSVSECKSGSERGSPIEKYVRAKHPNY 1080

QY 1081 SPGSPDIEKYQVYELFCPLFVHNDFOSEADWLRFWMSKYKLSVPGPHYLSHVPGLNPNQCN 1140
DB 1081 SPGSPDIEKYQVYELFCPLFVHNDFOSEADWLRFWMSKYKLSVPGPHYLSHVPGLNPNQCN 1140

QY 1141 YVPYPTFNLPFHSAVGSNDPLDLA:KHSRPGPTANGASKEKTKAPPNVKNEGP LNKV 1200
DB 1141 YVPYPTFNLPFHSAVGSNDPLDLA:KHSRPGPTANGASKEKTKAPPNVKNEGP LNKV 1200

QY 1201 KTKVDKRSQDELS:KCYHCGIVFDEVYALHMSCHDSGPFQCS:CHLCTDKYDETT 1260
DB 1201 KTKVDKRSQDELS:KCYHCGIVFDEVYALHMSCHDSGPFQCS:CHLCTDKYDETT 1260

QY 1261 HIQRLRNNAQVEKNKPKPE 1281
DB 1261 HIQRLRNNAQVEKNKPKPE 1281

RESULT 2

US-10-037-667-1
; Sequence 1, Application US/10017667
; Publication No. JS200202-77145A1
; GENERAL INFORMATION:
; APPLICANT: Motcar, Bruce A.
; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY
; TITLE OF INVENTION: LADDAJOS
; FILE REFERENCE: 10287-04400
; CURRENT APPLICATION NUMBER: US/10/037,667
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 60/243,110
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-667-1

Query Match 3.6%; Score 246; DB 14; Length 537;
Best Local Similarity 19.7%; Pred. No. 2.9e-10;
Matches 151; Conservative 82; Mismatches 235; Indels 338; Gaps 32;

QY 498 TDKSSSGAKKDFSKGAEDNMVTSYNCCFCDFRYSKSGPQ-----VTVVGPJLR 548
DB 29 TPNSQSSPSRSJANSIKVEM-----YSDESSRLGDPDERLLDXDSDVVEDSUSE 81

QY 549 HYQQLNHNKTIKHCPGPRGLC---SPERHLGEBITYPFACRKNCSHCAJLLHLSPG 605
DB 82 -----PLGYCDSGSGEPH-----SPG 97

QY 606 AAGSSRVKHQCHQCSFTTPDVLVLFVHESVHESQASDVQKQANHLQSDGQCV 655
DB 99 G:ELPGLKUKDCVGVXVIGIPVLMVHKRSHTGSPHFNCCGASFTQKGLRLHKLHLS 157

QY 656 GQGSVKESKESCTKCDFTQVBEELSHRYRAHSC-----YKRCQCSFTAACTOSL 707
DB 158 GEXPFK-----CPFNACRRDALTGHL-RTHSVSSFTVCKPKYCNVCGRSYKQOSTL 210

QY 708 LEHNTVTHCQEQDITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVAKR 767
DB 159 LEHNTVTHCQEQDITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVAKR 767

```

Db 211 EEKCHNYQLSLT-----DAQALT-----CQP-----235
QY 768 EKLEKCKLVWTESSDLDLNVWRGADILRGS---PSYTCASLGLTTPVSGTQECT 824
Db 236 -----GDBIRLEWVPSMLHPSTERTPTFIDRLANSUT-----268
QY 825 KTLRDSNVAAHLARPIYGLAVETKGFLOQAPAGGKSGALPQOYPASGENKSKDESQS 884
Db 269 -----KREKSTPOKF---VGEKQMRFLSD 290
QY 885 LRRRRGSGVFCANCLTTKTLRKNXNAGVVCNACG:YOKLHSTPLNIIKQNGEQ: 944
Db 291 -----PYDVNASGGY-----EKD 304
QY 945 IRRTRKRLNPEALQAEQLNQQRSGNEEQVNGSPLERRSEDLHTESHOREIPLPLSKY 1004
Db 305 VELVAHGLEPGF-----GSLAFVGTGHL-----APLRLP?-----336
QY 1005 EAQSLTKSHSAQOVLVVSOTLDIHKRMQPLHIQIKGPQESTGCPGNSSSVSEKGSSE 1064
Db 337 -----TNCISELTPVISS---VVTQMOPIPSRLELP-----GSRAGEGPELDGD 378
QY 1065 GSP:EKYMRPAKHNYSPGSP:PIEKYOYPLPGLJFVENDFQSEADWLRPWSKYKLVPCN 1124
Db 379 GGPL---LYRAGSLTDPGASP-----SNGQDSTC-----TESKH 410
QY 1125 PHYLSHVPELGN---PCQNVYVPTFNPPHPSAVGSDNDIPDLAIKHSRPGPTANGASK 1182
Db 412 EDRIQGVVSLQGGPPQ---PPPTIVGRHSPAVAKEDPKQE-GLLRGTGP?-----SK 462
QY 1183 EKTAPPNVKNEGLPNVVKTEKVRSTODELSTKCVHCGIVFLDEVYALHMSCHGDSGP 1242
Db 463 EVLRV-----VGESGPVKAFCHECHRIFLDHVMTFTHMGCHGRFP 505
QY 1243 FOCSCQHLCTDKYDFTTHIORGLHR 1268
Db 506 FECNICGHSQDRYEFSSHIVRGEK 531

RESULT 3
US-10-037-667-4
; Sequence 4, Application US/10037667
; Publication No. US20020177145A1
; GENERAL INFORMATION:
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY
; FILE REFERENCE: 10287-044001
; CURRENT APPLICATION NUMBER: US/10/037,667
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 60/243,110
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-667-4

Query Match 3.5%; Score 242; DB 14; Length 515;
Best Local Similarity 23.6%; Pred. No. 5,6e-10;
Matches 127; Conservative 66; Mismatches 199; Indels 146; Gaps 25;

QY 814 LTPVSGTQCTKTLRD-SPNVAAHLARPIYGLAVETKGFLOGAPA---GGEKSGALPQ 869
Db 37 LSTTSCAQNSKSDRGMASNVK-----VETQSDENGRACENGECAEDLRY 84
QY 870 YPASGENKS---KDESQSLRRRRSGV-----FCA-NCLTTKTLRKNXNAN 912
Db 85 LQASGEMNGMSHRDQSSAL-----SGVGIRLPNGK:KCDICGIVCIGPVNLVHVKRSH 119
QY 913 GG---YVCNACGL-----YOKLHSTPLNIIKQNGEQIIRRTKRLNPEALQ 959

```

```

Db 140 TGERPFOCNOGASFTQGNLRLRHILHSKEP---FKCHLCNVACRRR-----DALT 189
QY 960 AEQANKQQRGSNEQVNGVSPLEERS--EDHLTESHQ--REIPLPSL-----SKYEAQ 1007
Db 130 GHJRTSHVGFHKCYGCRYSYKORSSLEEHKEKCHNYLESMTGLFGVCPV:KEETNNKEMA 249
QY 1008 GSLTKSHSAQOVLVVSOTLDIHKRMQPLHIQIKSPQESTGCPGNSSSVSEKGSSESGSP 1367
Db 250 EDLCKI GAERSLVLDRLASNVAKRKSMM-----PQKFJGCKCLSDMPYDSANYEKEDMM 303
QY 1068 IEKYMRAKHP--NY-----SPQSG-----PFEKYQY-----PLGLFVNDPQ 1105
Db 304 TSHYDCAINNAIYGAESLRLPLVQTPGSSSEVVPVISSYQLHKPPSDGPPRSNESAQ 363
QY 1106 SEADWLRPWSKYKLSVPCNPHYLSHVPGLP-NPCQNVVP-----YPTNCLRP 1151
Db 364 DAVNLLLSKAK-SVS-----SERASPSNCCQDSTUTESNAEORSGLIYLTHNINP 416
QY 1152 HPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPRYKNEGPJNVVKTEKVDSTQD 1212
Db 417 R-----ARNGALJKEBQRAVEVLRAASENSQCAFVWSTSGE 453
QY 1212 ELST-KCVHCGIVFLDEVYALHMSCHGDSGPQCS-CQHLCTDKYDFTTHIORGLHR 1268
Db 454 QLVYKCECHVFLDHVMTFTHMGCHGRFPFECNMGYHSQDRYEFSSHIRGEHR 512

RESULT 4
US-10-205-823-321
; Sequence 321, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela Y.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 321
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-321

Query Match 3.5%; Score 240; DB 15; Length 419;
Best Local Similarity 25.7%; Pred. No. 5.7e-10;
Matches 119; Conservative 47; Mismatches 167; Indels 130; Gaps 25;

QY 853 LOGAPAGGKSGALPQOYPASGENKSKDESQSLRRRPGSGVFCANCLTTKTLRKNXNAN 912

```



```

Db 338 LKSEQRAYEVLRAASENQDAFRVWSTSGQLKVKCEHCRVLFDLHWYTHYGHSC 397
Qy 1241 G---PFCSTICQHLCTDKYDFTHIQRLHR 1268
Db 398 GFROPFECNMGYHSODRYEFSSHITRGHR 428

RESULT 9
US-09-764-864-947
; Sequence 947, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-947

```

```

Query Match 3.2%; Score 222; DB 10; Length 323;
Best Local Similarity 27.2%; Pred. No. 9.5e-09;
Matches 97; Conservative 38; Mismatches 128; Indels 94; Gaps 20;

Qy 367 QRGSEQVNGSPLEERSCHLTESHQREIPLPSLSKYEAQSLTKSHSAQGPV---LVS 1023
Db 8 ETGEKPHRCHLCPFASAYERHL-EAHR 48
Qy 1024 -----QTLDIHKMOPHLHIQKSPQESTGDPGNSSSVSEK--GSSRGGSP-EKYMRA 1075
Db 49 FRGSDRNLNLSHRRKHKVPIK-----GTRSLSSKKWGVLCVKTGNLGSRA 99
Qy 1076 KHPNYSPPGSP-EXYQYPLFGLPFVHNDFOSEADMLR---FMSKYKLSVPG----- 1123
Db 100 L-INLSPPSMVVKPDY-----LNDFTHEIPNICTDSYESMAKTTPTGGLPRCPOEL 150
Qy 1124 ---NP-HYLSHVPG-----LPNQCQ-----YVYPF---TFNL-PPHFSAVGS--DNDIPL 1164
Db 151 MVONPLNQLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOPSTQAVWSAVSASIPQ 210
Qy 1165 DLAIKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNVVKT-EKVDRTQD-EL 1213
Db 211 SSPTSPEPRPSHSQRYSPVAGSPSEPSAHTSTPSIGNSQP-----STPAPALPVQCDPQL 266
; US-09-764-864-947

```

```

US-09-764-864-1095
; Sequence 1095, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1095
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

US-09-764-864-1095
Query Match 3.2%; Score 222; DB 10; Length 323;
Best Local Similarity 27.2%; Pred. No. 9.5e-09;
Matches 97; Conservative 38; Mismatches 128; Indels 94; Gaps 20;

Qy 367 QRGSEQVNGSPLEERSCHLTESHQREIPLPSLSKYEAQSLTKSHSAQGPV---LVS 1023
Db 8 ETGEKPHRCHLCPFASAYERHL-EAHR 48
Qy 1024 -----QTLDIHKMOPHLHIQKSPQESTGDPGNSSSVSEK--GSSRGGSP-EKYMRA 1075
Db 49 FRGSDRNLNLSHRRKHKVPIK-----GTRSLSSKKWGVLCVKTGNLGSRA 99
Qy 1076 KHPNYSPPGSP-EXYQYPLFGLPFVHNDFOSEADMLR---FMSKYKLSVPG----- 1123
Db 100 L-INLSPPSMVVKPDY-----LNDFTHEIPNICTDSYESMAKTTPTGGLPRCPOEL 150
Qy 1124 ---NP-HYLSHVPG-----LPNQCQ-----YVYPF---TFNL-PPHFSAVGS--DNDIPL 1164
Db 151 MVONPLNQLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOPSTQAVWSAVSASIPQ 210
Qy 1165 DLAIKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNVVKT-EKVDRTQD-EL 1213
Db 211 SSPTSPEPRPSHSQRYSPVAGSPSEPSAHTSTPSIGNSQP-----STPAPALPVQCDPQL 266
; US-09-764-864-1095

```

```

RESULT 11
US-09-764-864-1515
; Sequence 1515, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1515
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1515

```

```

Query Match 3.2%; Score 217.5; DB 10; Length 294;
Best Local Similarity 28.5%; Pred. No. 1.8e-08;
Matches 39; Conservative 35; Mismatches 113; Indels 75; Gaps 18;

Qy 1012 KSHSAQGPV---LVS-----QTLDIHKMOPHLHIQKSPQESTGDPGNSSSVSEK--G 1060
Db 5 RSHTGEKPYKCELCSFRCSDRNLNLSHRRKHKVPIK-----GTRSLSSKKWVG 55
Qy 1061 SSERGSP-IEKYMRAKHPNYSPPGSP-IEKQYPLFGLPFVHNDFOSEADMLR---FMSKY 1117
Db 56 VLOKTSNLGYSRAA---NLSPSMVVMVQKQY-----LNDFTHEIPNICTDSYESMA 106
Qy 1118 KLSVPG-----NP-HYLSHVPG-----LPNQCQ-----YVYPF---TFNL-PP 1151
Db 107 KTTPTGGLPRDQELMVGNPLNQLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOP 166
Qy 1152 HFSAVGS--DNDIPLDLAIKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNV 1199
Db 167 STQAVWSAVSASIPQSSSTPSPEPRPSHSQRYSPVAGSPSEPSAHTSTPSIGNSQP--- 223

```

QY	1200	VKTEKVD	STQD	-ELSTKCV	HGTVFL	DEVMY	ALHMS	CHGDS	GPFGCS	ICOH	LCTK	YDF	1258
DB	224	-STPA	PAL	PQD	PQL	LHHC	QCD	MYFAD	NILYTH	MGCH	GYEN	PF	282
QY	1259	TTTH	QIRG	LHRN	1270								
DB	283	ACH	FARG	GOH	NOH	294							

RESULT 12

```

US-09-749-728B-11
:
: Sequence 11, Application: US/09749728B
: Patent No. US20020142457A1
: GENERAL INFORMATION:
: APPLICANT: Umezawa, Akihiro
: APPLICANT: Hata, Jun-ichi
: APPLICANT: Fukuda, Keiichi
: APPLICANT: Ogawa, Satoshi
: APPLICANT: Sakurada, Kazuhiro
: APPLICANT: Gojo, Satoshi
: APPLICANT: Yamada, Yoji
: TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARCINOMA
: FILE REFERENCE: 03766.000043
: CURRENT APPLICATION NUMBER: US/09/749,728B
: CURRENT FILING DATE: 2001-09-17
: PRIOR APPLICATION NUMBER: H11-372826
: PRIOR FILING DATE: 1999-12-28
: PRIOR APPLICATION NUMBER: PCT-JP00-01148
: PRIOR FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: PCT-JP00-07741
: PRIOR FILING DATE: 2000-11-02
: NUMBER OF SEQ ID NOS: 80
: SOFTWARE: PatentIn Ver.2.0
: SEQ ID NO 11
: LENGTH: 442
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-749-728B-11

```

Query Match	3.1%	Score 215;	DB 10;	Length 442;
Best Local Similarity	33.0%;	Pred. No. 5.6e-08;		
Matches 69;	Conservative 22;	Mismatches 68;	Indels 50;	Gaps 9;
Qy	873	SGENKSDSOSLLRRRGSGVFCANCLTTTISLRKQANGGYVCNACGLGYKLFSTPRP	932	
Db	248	NGINRPLIKPQRLLSASRRVCLSCANCOTTTTLRRNAESEPVCNACGLYMKLHGVRP	307	
Qy	933	LNII-KQNNGEQ-IRRRTRKRLNPEALQNKQOR-----GSNEEQVNGSPLERRSSDHL	988	
Db	308	LAMRKEG-----IQTRKRKPKN-----LNKSKTPAAPSGSSELPASGASSNSNAT	354	
Qy	989	TESHQREIPL---PSLSKYEACGLTKSHSAQQPVLVSQTL-----DIHKRWQFL	1035	
Db	355	TSSEENRNPKTFCGLSSH-----YCHSSS-----VSQTSVAMSMCHGSIHPVLISAL	403	
Qy	1036	HIQIKSQDSTGDPGNSSSVSEGGKSSSR	1064	
Db	404	KL---SPO-----GYASPVSQSPTSSK	423	

RES:11.T 13

```

RES001-13
US-09-974-298-114
Sequence 114, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10

```



```

DQ 687 SPVLPVGTNGS---RSSTPSPPLNLSSRNTQCYLYTAEGAEEQVPELDSLPKQ 743
QY 1101 HNDFOGEADWLRFWSKYKLSVPGNPHYLS-----HVPGLPNCQN 1140
DQ 744 QGELLERSTITSYQNSVSVQEPNLNLSCAKKEPKQKSCVTDSEBPNVNVIPPSANPINI 803
QY 1141 VVPYPTENLPPHESAVGSDNDIPDLAIKHSRPGPTANGASKKTKAPPNVK-----1192
DQ 804 APTVTAQPT-IVAADQNSVCLRAL-----AANKQILIPQVAYTSTVS 851
QY 1193 ---NEGFLNVV-----KTEKVDNSTQ-----DELST-----KCVHCG 1222
DQ 852 PAVOEPPPLKVIQBNQDERQDTSSEGVSNVEDQSDSTPPKKMKRKTENGMYACCLCD 911
QY 1222 IVDLDEVYALHMSCHGSDGPPQCSICQHLCTCKYDFTHIORGLHRNNAQVKNKG 1278
DQ 912 KIFQKSSLLRHKYEHTGRPHCEGICCKAFKHKKHLIEHMLHSGEKPYQCDKCK 968

RESULT 14
US-09-755-830-42
; Sequence 42, Application US/09755830
; Patent No. US20020104112A1
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Karia
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES
; FILE REFERENCE: 10287-047001
; CURRENT APPLICATION NUMBER: US/C9/755,830
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 38/283,300
; PRIOR FILING DATE: 1994-07-29
; PRIOR APPLICATION NUMBER: US 38/238,212
; PRIOR FILING DATE: 1994-05-02
; PRIOR APPLICATION NUMBER: US 08/121,438
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: US 07/946,233
; PRIOR FILING DATE: 1992-09-14
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-755-830-42

Query Match 3.1%; Score 210.5; DB 10; Length 376;
Best Local Similarity 22.3%; Pred. No. 9.6e-08;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

QY 814 LTPVSGTQETKTLRD-SPNVEAAHLARPIYGLAVETKGFJQCAPA---GGKSGALPQC 969
DQ 37 LSTTSQAQNSKSDRGMAENVK-----VETQSDENGACENMGEECAEDLRM 84
QY 870 YPASGENKS---KDESQSLRRRRGSGVFCANCLTKTSLMRKNANGGVVQACGLYCKL 926
DQ 85 LDASGEQWNSHRDQGSAL-----SGVGI-----RLPMGLKLCDCIGIV---125
QY 927 HSTPRPLNIKQNGEOTIIRRTKRLNPEALQAEQLNQKQSGNEEQVNGSPLERRSDD 986
DQ 126 -----CIGPNVLM-----VHKRS--138
QY 987 HLTESHOREITPLSLKYAAGSLTKSHSAQOPVLVSQTLDIHKRQPHLIQIKPQEST 1046
DQ 139 HTGDKCLSDMPYDS-ANYEKEDMMT-SHVMDQDINNAINYLGAESURPL---VQTP-189
QY 1047 GDPNGSSSVSEGGKSGSRGSPFIEKYMPAPAKHPNYPGSPFIEKYQVPLGLPFVHDFOS 1106
DQ 190 -----PGSEVVPVSSMYQLHKPDSGDPRS-----NHSAD 222
QY 1107 EADWLRFWSKYKLSVPGNPHYLSVPGLP-NPCQNYVP-----YPTNLPPH 1152
DQ 223 AVDNLLLLSKAK-SVS-----SREASPSNSQDSTTESNAEQRSGLVLTNHNPH 275

```

```

QY 1153 FSAVGSNDIPDLAIKHSRPGPTANGASKKTKAPPNVKNEGPJNVVKIEKVDRTSCDE 1212
DQ 276 -----ARNGLALKEBQRAYEVLRAASENSCCAFRVVSTSGEQ 312
QY 1213 LGSF-KCVHCGIVFLDEVYALHMSCHGSDG---PQCSICQHLCTCKYDFTHIORGLHR 1268
DQ 313 LKVKYKCEHCVFLDHWMTTHKGGCHGCHGRDPEFCNMCGYHSCDRYEFSSHITRSEHR 372

RESULT 15
US-10-177-293-174
; Sequence 174, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Giatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Yareen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Roersch, Sebastian
; APPLICANT: Konahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMEBITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-174

Query Match 3.1%; Score 209; DB 15; Length 474;
Best Local Similarity 26.4%; Pred. No. 1.8e-07;
Matches 97; Conservative 33; Mismatches 107; Indels 130; Gaps 19;

QY 750 LTPD-----SKMGEPVSESV-VKREKLEEKDGLK-EKWTFESSDDJRNVTWKGADILRSP 804
DQ 175 LSPDPSTTGAASPASSAGSGARGEDKGVKYQSLTESYKME-----SGRP 222
QY 805 SYTQASLGLLTPVSGTQETKTLRDSPNVEAAHLARPIYGLAVETK-----FLOCAP 857
DQ 223 -----LRPGLAT--MGIQ-----PATHHPITPYPSVVPAAAHIDYSSGLFHPGSF 264
QY 858 AGGKSGALPQC-----YPASGENK 877
DQ 265 LGGPASSFTPKREKTRSCSEGRECVNCGATATPLWRBDGTGHI:CNACGFVHKMGQNR 324

```

```

QY      878 SKDESOLLRRRRGSGVFCANCLTTKTSLWRKNVANGYVCNACGLYOKLHSTPRPLNKK 937
Db      325 PLIKPRLSAAARRAGTCCANCOTTTTLWRRNANGDPVCNACGLYKLNVRPETYKK 384
QY      938 QNNGEQIIRRTTKRLNPEALQAEQLNKOQGSN--EE-----QVNGSPLEERS-EDHLT 989
Db      385 EG-----IOTNRKMSN-----KSKKKGKGAECFELSCKMQEKSPPSAAALAGHMA 432
QY      990 ESHQREIPLPSLSKYEAQGSLLTKSHSAQOPVLVSOTLD-HKMQPLHIQIKSPQESTGDP 1049
Db      433 -----PMGHLPPF-----SHSGH--LEPTPT-----PIH---PSSLSFGHP 464
QY      1050 GNSSSVS 1056
Db      465 HPSSMVT 471

```

Search completed: October 29, 2003, 11:53:09
 Job time : 232 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:57 ; Search time 72 Seconds

(without alignments)
1711.001 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKNPPLRNVAEGEQI.....IQRGLHRNNAQVENKGFKE :281

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	4.0	1186	2 T33754	O/E-1-associated z
2	268	3.9	758	2 A29253	finger protein hun
3	248.5	3.6	1114	2 I50222	delta2 - chicken
4	241	3.5	452	2 C41602	transcription fact
5	239.5	3.5	518	2 A56355	DNA-binding protei
6	234	3.4	782	2 S04047	finger protein zfy
7	233	3.4	1350	2 S00647	finger protein - A
8	222.5	3.2	466	2 A36389	transcription fact
9	222.5	3.2	982	2 T43676	hunchback-related
10	218.5	3.2	816	2 S05548	gap protein hunchb
11	218	3.2	783	2 A31491	sex-determining re
12	215	3.1	439	2 I57561	transcription fact
13	215	3.1	480	2 A40815	transcription fact
14	213	3.1	392	2 A56229	lymphoid transcrip
15	213	3.1	779	1 S40382	box A-binding fact
16	212.5	3.1	440	2 I61183	transcription fact
17	212	3.1	431	2 I59572	Ikaros DNA binding
18	209.5	3.1	444	1 A39794	transcription fact
19	209	3.1	474	2 A41782	transcription fact
20	209	3.1	2783	1 A41948	alpha-fetoprotein
21	207.5	3.0	444	2 B36389	transcription fact
22	207	3.0	443	2 B39794	transcription fact
23	207	3.0	1124	2 JX0293	zinc finger protei
24	207	3.0	1828	2 A40115	microtubule-associ
25	206.5	3.0	390	2 I54119	transcription fact
26	206	3.0	427	2 B62229	lymphoid transcrip
27	205.5	3.0	439	1 A48099	transcription fact
28	205.5	3.0	1036	1 A34755	nitrogen regulator
29	205	3.0	380	2 I50701	transcription fact

transcription fact
transcription fact
GATA-type transcr
transcription fact
eyeless, long form
URSL protein - sm
E-box-binding rep
zfa protein - mous
finger protein 2FY
transcription fact
transcription fact
homeotic protein 2
regulatory protein
cell proliferation
transcription fact
microtubule-associ

30 205 3.0 698 2 S25409
31 203.5 3.0 486 1 A57601
32 203 3.0 564 2 T38291
33 202.5 3.0 391 2 I50702
34 202.5 3.0 838 2 I45557
35 200.5 2.9 950 2 S27473
36 200 2.9 154 2 A56242
37 199.5 2.9 742 2 S12533
38 199.5 2.9 754 2 S26648
39 198.5 2.9 388 2 I51420
40 197 2.9 564 2 T43298
41 196.5 2.9 305 2 S33642
42 195 2.8 876 1 A57988
43 194.5 2.8 2938 2 T30249
44 193.5 2.8 435 2 D41602
45 193.5 2.8 1825 2 S13507

ALIGNMENTS

RESULT :
T33754
O/E-1-associated zinc finger protein Roaz - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 02-Sep-2003 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T33754
R/Teal, R.V.L.; Reed, R.R.
submitted to the EMBL Data Library, March 1997
A/Description: Cloning and functional characterization of Roaz, a zinc finger protein t
A/Reference number: Z21399
A/Accession: T33754
A/Status: Preliminary; translated from GB/EMBL/DBCB
A/Molecule type: mRNA
A/Residues: 1-1186 <TSA>
A/Cross-references: EMBL:J52564; NID:G2149791; PID:G2149792; PIDN:AAB58646.;
A/Experimental source: strain Sprague-Dawley
C/Function:
A/Description: regulates the temporal and spatial pattern of olfactory neuronal-specif

Query Match 4.0%; Score 271; DB 2; Length 1186;
Best Local Similarity 20.0%; Pred No 13e-07;
Matches 275; Conservative 156; Mismatches 515; Indels 432; Gaps 69;

QY 52 AELNHKEEHLHVQDPSSSSKKDLKSAVLSEKAGFNYESFSGKGNFPSPFHEVDTPDKXL 111
DB 84 AAFSRADHLKHLKTHSS--KPFKCSVC--KRGFSSTSLQS-----HMCARKKNK- 131
QY 112 AFSFPAAGGVCEPL-KSPQRAEDDPQDMACTPSSDGL-ETKEDQK----VSPKATEETG 165
DB 132 -----EHLAKSEAKKD---DFMCDYCEDTFSTCELEKHVJTLHPQUSEKAD 177
QY 166 -----QAQSCQANCOGLSPVSVA-----SKNPQ 188
DB 178 LQCIHCPEVVDSTLAHHQAHANQKHKCPMCPEQFSSVEGVYCHLDHSHRQDPSSNHS 237
QY 189 VPSD---GGVRLNKSTDLLVNDNPD-----AP-SPELQCFKCNICGYGYGNPDTDJ 240
DB 238 VSPDPVLGSAVSSATP---DSTDPVLGSAVSSATPSSASV-----ERGSTPDSTL 290
QY 241 KHPRKVHLGHNHRTQDAEJDSKILALHNMVAFSHSKQFQKYNRSVFGSVLQDINSSRPV 300
DB 292 KPIRG-----QKKMRDGGQSWSKV---YSPCYCKRCDTSJ--AVLEIHEKATINADKP- 339
QY 301 LKNGTYDVQVTSGGTFIG'G---RKT-----PDQCGNTKYFRCKFCNFTYMGNSSTEL 350
DB 340 QQSHTQCIQCLDSMPTLYLNENHVRKLKSHAPVWQFGNISAFHCNYCPDMFADINS--J 397
QY 351 EQHFLOTH--PNKIKASLFSSESAKPSKXS----NKSIPALQSSDSGDLGKMQCKITVK 404
DB 398 QEHIRVSHCGPN-----ANPPDGNNAFFCNQCSMGFLTSS--LTEHIQACHCS 444
QY 405 AGD---DTPVGISVP;KPLDSSRQNGTEATSYNCKFKCSFSCSESSSLKLLSHYKQHQGA 462

Db 674 NGNSSSSNGTTSVAAPPSTPAAGAIYEKCYCDIFFKDAVLVLTIMHYSQDVF 733
 QY 1244 OCSICCHLCTDKYDFTTHIQRLH 1267
 Db 734 KNCMGEKCDGPGVLVFMARNAH 757

RESULT 3

i50222
 deltraEF1 - Chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 2i-Jul-2000
 R:Accession: I50222
 R:Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.
 Development 119, 433-446, 1993
 A:Title: Delta-crystallin enhancer binding protein delta EBF1 is a zinc finger homeodomain
 A:Reference number: I50222; MUID:94116444; PMID:7904558
 A:Accession: I50222
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-114 <FUN>
 A:Cross-references: GB:D14313; NID:g391631; PIDN:BAAC3259.1; PID:g391632

Query Match 3.6%; Score 248.5; DB 2; Length 114;
 Best Local Similarity 18.8%; Pred. No. 2.4e-06;
 Matches 210; Conservative 144; Mismatches 377; Indels 385; Gaps 52;

QY 378 NSNGIPALQSSDGLQKW-----QDKITVRAG-----DTPGVGVPIKPLDSSRQN 426

Db 22 NYNVEIA--NSDSODEKLHIVEESITDAACDASVPEDDLPTDHTV--LFENESREG 77

QY 427 GTATSYWCKFCSPSCSSSSSLK--LEHYGKHGAQVQSGGLNPELNKLRGSGVINONDL 486

Db 78 S* N-----SC-----WEDEGKETKEI-----LGPE----- 97

QY 487 AKSSGEITWTKDKSSGAKKDFSSKGAEDNMVTSYNCQCFDFRYKSHGPDVIVGPL 546

Db 98 AQSDVEGCTVKEDECDSDAENQNDHPNVEFL-----QEDTAVTYPE 141

QY 547 LRHYQQLHNIHKCT-----INHCPCPRG---LCSPEKHGEITYPFACRKS 590

Db 142 APEEDORQGTPEASGDENGTPDAFSQLLTCPCYCDRGYKRTSLKEH--KYRHEKNEED 198

QY 591 NCSHCALLLHLSPGAAGSSRVKHQCHQCSFTTPDVVLLPHYESVHESQADYK--QEA 648

Db 199 NFS-----CSLCSYT-----FAIRQLDRHMTSHKSGRQD 228

QY 649 NHLQSDGQGSVKESKESCTKCDFTITQVEEISRYHRAHS---CYKCRQCSFTAAQTQ 705

Db 229 RHVTQSSGNRKFK-----CTEGKAFKYHHLKEHL--RIHSGEKPYCPCNCKKXFSHSG 281

QY 706 SLEHPTVHCOEQDITANGBEDGHAISTIKBPKIDFRVYNLLTP--DSKMGEPVSES 763

Db 282 SYSSHISKKK-----IGLMPVKGRARSGLKTSQCSSPSLSASPSPARPQ 327

QY 764 VVKREKLEEKDGLKE-----KYWTE-----SSDDLNVTVWRGADILRGS 803

Db 328 I--RQIKENKP-LQEGLPVNIKTEPVYEFKPIVVASINGNSTPQNGVFGSGPLQAT 384

QY 804 PS---YTOA---SLGLLTPVS-----GTOEQTKT 826

Db 385 SSPQGVQAVLPTVGLVSPISNLSIDIONLVKVAVDGNVPCVLENNEANLASKEQ-ET 443

QY 827 LRDSPNVEAAH-----LAPPIVGLAVETKGFLOGAPAGEKSGAJPQOYP 871

Db 444 ISNASTQAGHSLISAISLPLVDQDQTKIINYSLEQPSQLQVVPQNLKHSYV----- 499

QY 872 ASGENSKDSQSLRRRSGSGVFCANCLTTKSLWRKNANG-----YVNCACGLYCK 925

Db 500 ---TNSCKNEK-----LPEDLTIVSEK--DRNFEGETNDS--CLLCDDC----- 537

QY 926 LHSTPRPLNLIKONNGEQIIRRTTRKRLNPEALQABQLNKQQRGSNEEQVNSPLERRSE 985

Db 538 ----POCLNALQE-----LKHVETK--NPPQL-----POSSGTEAEKPPSPAPSETGE 579
 QY 986 DHLTESHOREIPLPS-LSKYBAQGLTKSHSAQCPVLVSQT-----LQTHK-----RMOPLH 1036
 Db 580 NNLSPG---QPP-LKVLJLLKAYYALNAQPSAEELSADIADSVNLPJDVVKWFEKQAGC 636
 QY 1037 IQIKS-----POES*GDGNGSSSVSEK-----GSSER 1064
 Db 637 ISVQSSGPPSPQVXKISSPTNQDQAATTNESEPQSNNSQNPANTSKSQTSSGGSTQN 696
 QY 1065 GSPIEKYMRPAKHP-----NYSPPGSPTEKVOY?--FCLPVHNDQFSEADWL 1111
 Db 697 GS---RSSTPSPFNLSSRSNSQYTYTAEQVQEPQMEPLDLSLPKQFGLLERSTIT 753
 QY 1112 RWSKYKLSVPGNPHVL-----SHVFGLPNCPQNYVPYPTFNLP 1151
 Db 754 SVYQNSVYVQSEFPLNLCAKKEPKENSDITDSDFVNVIFPSANPINAIPTVTAQLPT 813
 QY 1152 HFSVQSGNDIP-----LQLA-KHSRPGPT-ANGASKK-- 1184
 Db 814 -TVAIDQNSVPCLRALAANKQTI-LPCVAYTYSTTVFAVQETPPKQTQANGSQDERQD 872
 QY 1165 --TKAPPNVKNEPFLNVKTEKVDSTODELSTKCVHCGIVFLDEVVYALHMSCHGDSQP 1242
 Db 873 TSSGVSQVWEDQNDSDS-PPKXKXKRTKNGYA-CDLCKIFQKSSSLRKHYEHTGKAP 931
 QY 1243 FQCS*CGHLCTDKYFTTHIQRLHNNNAQVEKCK 1278
 Db 932 HEGGICKAKFKKHH--EHRM-LHSGEKPYQCDKCK 967

RESULT 4

C41602

transcription factor GATA-2 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 26-Aug-1999

C:Accession: C41602

R:Zot, L.F.; Mather, C.; Burgess, S.; Boice, M.E.; Harland, R.M.; Crkin, S.H.

Proc. Natl. Acad. Sci. U.S.A. 88, 10642-10646, 1991

A:Title: Expression of GATA-binding proteins during embryonic development in Xenopus 1

A:Reference number: A41602; MUID:92073343; PMID:1961730

A:Accession: C41602

A:Molecule type: mRNA

A:Residues: 1-452 <ZON>

A:Cross-references: GB:N76564; NID:9214170; PIDN:AAA49723.1; PID:9214171

C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology

C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger

F:264-317/Domain: GATA-type zinc finger homology <GZF1>

F:267-291/Region: zinc finger GATA motif

F:318-371/Domain: GATA-type zinc finger homology <GZF2>

F:321-345/Region: zinc finger GATA motif

Query Match 3.5%; Score 241; DB 2; Length 452;

Best Local Similarity 27.8%; Pred. No. 1.8e-06;

Matches 99; Conservative 39; Mismatches 100; Indels 118; Gaps 19;

QY 750 LTPDSXGEPVSESVVKREKLEKQCKLKEKWTSSDOLRNVTWRGACILRGSPSYTCA 809

Db 163 VSPDQSPGAPFPSS-----RLCKOSIKYQM---SLSEGMK---MEGSPPLRSS 235

QY 810 SLGLLTPVSGTQEQT-KTLRDSFN-VEAAHLARP-YGLAVETKGFLOGAPAGEKSGALP 867

Db 206 ----LAPM-GTQCS:HP:PTYSVVPAAH-----DYSSGLFHPGSLGCPA----SSSTP 252

QY 968 QQ-----YPASGENKSKDESQSLLR 887

Db 253 KQSKSSECSSEGCVCNCGATATPLWRDGTGHYLCNACGLYKXNGQNRPLKEKRSL 312

QY 988 RRGSGVFCANCLTTKSLWRKNANGGVVNCAGLYQKLHSTPRPLNTIKQNGSQIRK 947

Db 313 AAPRAGTCCANCTSTTLWRNRNAGCPVCNACGYYKLHKNRPLPTKKKEG-----CT 367

QY 946 RTKELNPEALQAEQLNKQQRGSN-EE-----QVNGSPLEERSSEDLHTESHQREIPLPS 1000

```

Db      368  RNKMSN-----KSKKKKSGCEFEUSRCMQKSPFSA--LASHM--APNGH 414
Qy      1001  LSKYEAQGSITKSHSAQPVLSQTLIHKRMQPLHIQIKSPQESTGDPGNSSSVS 1056
Db      415  LAPPFHSQ---HILQTP-----PIH---PSSSLFGHPHSSXVT 449

RESULT 5
A56355
DNA-binding protein Ikaros form 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 01-Dec-2000
C:Accession: A56355
R:Moinar, A.; Georgopoulos, K.
Mol. Cell. Biol. 14, 8292-8303, 1994
A:Title: The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins
A:Reference number: A56355; MUID:95059058; PMID:7969165
A:Accession: A56355
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-518 <MOL>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: alternative splicing; DNA binding

Query Match      3.5%; Score 239.5; DB 2; Length 518;
Best Local Similarity 23.7%; Pred. No. 2.7e-06;
Matches 128; Conservative 66; Mismatches 198; Indels 149; Gaps 26;

Qy      814  LTPVSGTQBTKLKD-SPNVEAAHLAPPIYGLAVETKGFLOQAF---GGEKSAIAFQ 869
Db      37  LSTTSQAQNSKSDRGASNVK-----VETQSEENGACEMNGEACEADLRM 84

Qy      870  YPASGENKS---KDESQSLRRRGSGV-----FCA-NCLTTK-TSLARKYAK 912
Db      85  LDASGEKNGSHRDQSSAL-----SGVGG-RLPNGKLKCD-CGIVCIGPVLMVHKESH 139

Qy      913  GG---YVNCAGL-----YQLHSTPRPLNIKKNGGEIIRRTKRLNPEALQ 959
Db      140  TGERPFECNQGASFTQGNLRLHIKLSGEXP---FKCHLCNYACRRP-----DALT 189

Qy      960  AEQLNKQORGSGNEROVNGSPLEERS--EDHLTESHQ--REIPLPSL-----SKYBAQ 1007
Db      190  GHLRTHSVGPHKCYCGRSYKORSSLEHKEKCHNYLESMLPGMYPIKETHNEVA 249

Qy      1008  GSITKSHSAQPVLSQTLIHKRMQPLHIQIKSPQESTGDPGNSSSVSEKGSSEKSP 1067
Db      250  EDLCIKGAERSLVLDRLASNAVKRKSMM-----PQKFLGCKLSDMPYDSANYEKEDMM 303

Qy      1068  IEKYMRAKHP--NY-----SPRGS-----PIEKYQY-----PLFGLPFVHNDQ 1105
Db      304  TSHVMDQAINAINYLGAESLRPLVQTTPGSGSEVVPVSSYQLHKPPSGDPPRSHSAQ 363

Qy      1106  SEADMLRFWSKYKLSVPGNPHYLSHVPLG-PNCONYVP-----YPTFNLP 1151
Db      364  DAVDNLLLSKAK-SVS-----SEREASPSNQDSDTESNAEQRSGLIYLTNHNP 416

Qy      1152  HFSAVGSDNDIPLDLAIKHSRPGPTANGASKETKAPPNVKIEGP-VNVVTEKVDNRSTQD 1211
Db      417  H-----ARNGLAKBEQRAYEVLRAASENSODAFRVYSTGE 453

Qy      1212  ELST-KCVHCGIVFLDEVMYALHMSCHDSG---PFQCSICQHLCTDKYDFTTHIQRLH 1267
Db      454  QLKVYKCEHCVFLDHWMTYIHMGCHGCHGRDPFCNMGCVHSODRVEFSSHITRGEH 513

Qy      1268  R 1268
Db      514  R 514

RESULT 6
S24047
finger protein zfy-1 - mouse

```

```

N:Alternate names: Probable testis-determining factor
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S04047
R:Ashworth, A.; Swift, S.; Afara, N.
Nucleic Acids Res. 17, 2864, 1989
A:Title: Sequence of cDNA for murine zfy-1, a candidate for Tdy.
A:Reference number: S04047; MUID:89240049; PMID:2497440
A:Accession: S04047
A:Molecule type: mRNA
A:Residues: 1-782 <ASH>
A:Cross-references: EMBL:X14382; NID:955478; PIDN:CAA32552.1; PID:955479
C:Genetics:
A:Gene: zfy-1
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match      3.4%; Score 234; DB 2; Length 782;
Best Local Similarity 18.8%; Pred. No. 1e-05;
Matches 146; Conservative 93; Mismatches 258; Indels 278; Gaps 34;

Qy      20  LLEPIGTESKYSKVKESFSAQMSGENTDQSDAAELNKKHEHSLHVQPPSSSKKD----- 74
Db      179  VLDSGMPLEQCNDRKNCEDYLXMSLDEPSKTDLEGSSEVTNKAESSTGSKLCEASPE 238

Qy      75  LKSAVLSEKAGFN-----YESPSKGGNFPFPHDEVTRNMLAFSPFAAGGVCPEL 125
Db      239  VIKVCILKADSEVDVGTIHAVERSEKNGK-----EAEVTDQS-TSIRVPRVNVYSAS 292

Qy      126  KSPQRAE-----ADCFQ-----DVACTSPSGSLSTK-EDQ 154
Db      293  DSKSEEDTEVTVGDEADAGTAUTPEHEQOMVSEIKAAFLPAWTAAYENNSDE-EDQ 352

Qy      155  KMSPKATEETGCAQGGQANCCQLSPV-----SVASKNQVFSQGVRLNKSKTDLVNDP 210
Db      353  NVTASAL-LKQGES-----GLDVPKQSKKKRPS-----SKCYKSAIFV----- 393

Qy      211  DPAPISPELODFKNCVGYGYGNDPTDLIKHFKYHLGLH-----NRTQDAELD--S 262
Db      394  --APGGQT-UVVPCVFCGKFF-----KTKRFLKRHTKHPKPEYLANKKYHCTECDYSTN 444

Qy      263  KILALHNVQFESHKDFOKVNRSVFSGVLODINSRPPVLLNGTYDVQVTSGETFIGR 322
Db      445  KKLISLHNYE-SKLITKTEKTECCCKNLSPAGLCTHKTYTE----- 490

Qy      323  TPCQGNKYFRCKFNFTYMGNSSTELSQHPLQHPNKKIKASLPSEVAKPSEKNKS 382
Db      491  ---KGNKTKCKKFCD--YETAEQTLNHHLLVYH-----RKK 523

Qy      383  IPALQSSDSDGLGK-----WCKKITVKAGDDTPGVSVPIKP-DSRRONGTEATSYW 435
Db      524  FPHI-----CGEGKGFRRPSALKKHIRVHTGE-----KP-----YE 555

Qy      436  CKFCSPSCSSSLKLEHYGKHQGAQVSGGLNPGLNPELNDKLSRGSVINONDLAKSSE 495
Db      556  CQYCEYKSAQSSNLK--THIKSKH----- 577

Qy      496  TKTKDSSSGAKKSDS-SKGAEDNMV-----TSYNCFQDFRYSKSHGPDV-VVVGPLLR 549
Db      578  SKELPKKGCILLTTSOTKEAQOHAVLHQRSTHOCCHCNHKSNS-----SKAKRH 629

Qy      550  YQGLNHNHKTIKHPFPFPRGLCSP---EKHLGEITYTFACRKNCGHCALLLHLSPGA 606
Db      630  IISVHT--KAYPHKDCMCKSGFRRSELKQHV-----A 660

Qy      637  AGSSRVKHQCHQCSFTTPDPVLLPHYESVHESQASCVKQEAHLQSDGQGSQSVKESKEH 666
Db      662  THSKXKHQCHQCHDENSEDPFLSHHLSAH-----TKKVPF 697

Qy      667  SCTKCDFITQVEEISRHVRRAS---CYKRCQCSFTAAADTCSLLEHENTVHCCE 718
Db      698  KCKRCKKEFCQCELCQTH-YKTHSSRKVYQCEYCEYSTKQASGFKRHVTS-HTKD 751

```

```

RESULT 7
S00647
finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C:Accession: S00647
R:Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
A:Reference number: S00647; PMID:88082679; PMID:2826129
A:Accession: S00647
A:Molecule type: mRNA
A:Residues: 1-1350 <RUI>
A:Cross-references: EMBL:X06021
A:Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiation site.
A:Genetics:
A:Gene: fin
C:Keywords: DNA binding; zinc finger

Query Match      3.4%; Score 233; DB 2; Length 1350;
Best Local Similarity 19.3%; Pred. No. 2.5e-05;
Matches 239; Conservative 138; Mismatches 39; Indels 472; Gaps 70;

QY 222 FKCNICGYGVNDTDLIKHPRKYLHGHNRTRODAELDSKILALHAKWQF----- 273
DB 354 YLCSHCNKGFGQN--SDLVGHFRT--HTG--ERPQCAECHKGFIQKSLDLVHJLHRTHTGK 408
QY 274 ---SHSKDFQKVRNSVFGVLQDINSRPVLLNGTYDVQVTSGGTFGIGRKTPDCC-- 326
DB 409 PFKCSHC-DKKFETSAIAKHQTHTGKPY-----KSCDCKE 446
QY 327 -----QGNTRYKCFKCNFTYMGNSSTLEQHLQTHPNKIKASLPSSVAK 373
DB 447 FTQRNLIHQHRTHTGERPYKCTLCDRTFIQNS--DLVKH-----QKVHANLPLSD-- 495
QY 374 PSEKNS-----NKSIPALQS-----SDSGDGLKKQDKITV 403
DB 496 PHTANSPHKCKSCDUTFSHWSTFMKSLHSGEKFKQCAECKKGFTQKSDLVK--HHRV 552
QY 404 KAGDTPGVGYPIKPL---DSRQNGTEATSYW-----CKFCFSCESSSLK--LL 452
DB 553 HTGEK-----PFKCLLCKSPQNSD--LHKHWRHTGKFPFCYCDXSTFERSALI 603
QY 453 EHYGKHQGVQGGINPELNDKLSRGSVINQDLAKSBEGETMTKTDKSSGAKKDFSS 512
DB 604 KHRHRTHTG-----ERPHKCSVCQKGIQKSA--LTKHSTRHTGK----- 641
QY 513 KQAEONNVTSYNCFDFRYSKSHGPDVIVVQPLRHVQOLHNIHK---CT----- 560
DB 642 -----PYPTQC-----GKSPFQNSDLVKH-ORHITGSKPYHCTECNKRFTG 693
QY 561 ---IKH-----CPFCPRGLCSPE---KHL-----GE-----ITYPFACRKS-- 591
DB 684 SSVLVKRRHTHSGEKPYRCQCEKTFIQSSDLVYH-LVWNGENPPATAFHEILIERNLT 743
QY 592 -----CSHCA-----LILLHLSPGAAGSRVKHQHQCQS---FTTPDQVVLPHY 633
DB 744 RSEPDYPCTEGKGVHQRPALLKHL---RTHKTEKRYPCNECDKSFQTSQD---LVKH 796
QY 634 ESVHESQASDVQKEAN--HLQSDGQOSVKESKEH-----SCTKCD--FTQVSEELSR 683
DB 797 LRHTHTGERPYHCECNKGFIQNSD---LVKHQHTHTGERPYTCSQCDKGFQIR--SALT 851
QY 684 HYR---RAHSCYKRCOSFTTAADTQSLLEHFN--VHCOEQDITTANGEEQDHA-STIKER 741
DB 852 HNRHTHTGERPYKCEQCKCFIQNSDLVKH-QRIH-----TGKPYHC----- 892
QY 742 IDFRVYNLLTPDSKMGEPVSESVVREKLEEK-----DGLKEKWTSESSDDLNRVWFG 796
DB 893 -----PDCDKRFTEGSSLIKHQRTHSRNIKPYECGVCKGKSPSQSS----- 931
QY 797 ADILRSPSYTQASGLLTLPVSGTQE-----QTKTLRDSPNVEAAHLARPIYGLAVETKG 851

```

```

DB 932 -NLLKHLKCHSEQN---PPVALSSSELGFVACTQTHPDPVQ---HI---VYG---DTAS 976
QY 652 FLQGAPAGGKSGALPQCYPASGENKSKDESQSLPRRRGSGVFCANCLTTKTSLWKNXA 911
DB 977 YI-SPEAAERS----- 987
QY 912 NGGYVCNACS-----LYOKLHSTPRPLNIKQNGEQ:IRRTKRKLNPEALQAE 961
DB 988 ---FKCNDGCKCFAPRSVLIKHVRHTGERPKYC-----SQCTRSFIQKS 1029
QY 962 CLNKQKQSGNEECVNGSPLEERS--EDHLTESHQREIPLPSLSKYEAQCSLT:KSHSAQGP 1019
DB 1030 DLYKHRTHTHTGERPKYKGLCERSFVKSALSQR-----VHKVEEP 1071
QY 1020 VLYSQDLTIHKMCPJHIQIKSFOESTGPGNSSSVSEKGSSEKSGSP--IEKYVRPAKH 1077
DB 1072 VL-----NSAME--CCQVYWGSEKCDP--NSLPQQLHVIKEBESPH:VNAYSPLSL 1120
QY 1078 PNYSPF-----GSPTEKYOVYELFGLPFVHNDQFSEADWLRFMSKYKLSVPGNPHYLSHVP 1132
DB 1121 QSYFPPILPEKGTG--RYSCECEGCKCFTHKSV-----FJKHWR-----VHTG 1160
QY 1133 GLNPFQ-----NVPYPTFFNLPHFSVAGSDNDIP:DLAIKHSRPGGTANG 1179
DB 1161 EQPYTCCKEKGKSPSQSALVKKHVRHTGKPYACSTCGKSP:QKSDLA-KHOR-----HT 1215
QY 1180 ASKEKTKAPFNKVEGPNLVKTEKVDRTSTQDELSTKCVHCGIVFJDEVVYALHMSCHGD 1239
DB 1216 GERPYCTVCGKFI:DRSSVVKH---SR:HTGERPKYKNECT:KGFVCKSDLVKHMRTHTG 1272
QY 1240 SGFFQCSICQHLCTDKYDFTTHICRGLHNNNAQVENKNGKP 1279
DB 1273 EKPYGNCDDR-----SESTHSASVRHQ---RMCNTGRP 1303

RESULT 8
A36389
transcription factor NF-E1b - chicken
N:Alternate names: DNA-binding protein: NF-E1b
C:Species: Gallus gallus (chicken)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 26-Aug-1999
C:Accession: A36389; S15645
R:Yamamoto, M.; Ko, L.S.; Leonard, M.W.; Beug, H.; Orkin, S.H.; Engel, J.D.
Genes Dev. 4, 1650-1662, 1990
A:Title: Activity and tissue-specific expression of the transcription factor NF-E1 mul
A:Reference number: A36389; PMID:91065513; PMID:2249770
A:Accession: A36389
A:Molecule type: mRNA
A:Residues: 1-466 <YAM>
A:Cross-references: EMBL:X56930; NID:G62965; PIDD:CAA40952.1; PIDD:G62966
C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
P:279-311/Domain: GATA-type zinc finger homology <GZFI>
P:261-305/Region: zinc finger GATA motif
P:332-385/Domain: GATA-type zinc finger homology <GZP2>
P:335-359/Region: zinc finger GATA motif

Query Match      3.2%; Score 222.5; DB 2; Length 466;
Best Local Similarity 28.1%; Pred. No. 2.3e-05;
Matches 93; Conservative 37; Mismatches 138; Indels 63; Gaps 14;

QY 750 LTPDSKMGEPVSSVVKREKLEKQGLK-----EKWTSESSDDLNRVWTKGAD----- 798
DB 172 VSPDPNSTSAASPSSAGARQEDKQSIKYCVS:SEGKMKMESAPLSRLTSMGAQCESTHS 231
QY 799 ILGSSSYTQA-----SLGLLTPVS--GTQETKTLRDSNVVEAAHLARPIYGLAVETKGF 852
DB 232 PIPTYFSYVFAAHDYSSSJFHPGTFJGGPASSFTPRSKARSCSREGCVKNGCATATPL 291
QY 853 LOGAPAGGKSGALPQCYPASGENKSKDESQSLRRRRSGYFCANCLTTKTSLWKNAN 912
DB 292 WRDGTGTHYLCNACGLYHKXNGQNRILKPRPLSAAARRAGTCCACCCTTTTLWRNAN 951

```

```

Db      474 SFGVYALDQKAISENPSFPEGGI::NLASALGVVSNAIKGCTP::SPBKQSN 524
Qy      579 GBITYPFACRKSNCSHCALLHLHLSPGAAGSSRYK-HQHCCSFITPDVDVLHFHYESYH 637
Db      525 GE-----CRSS::SGKIKFKCKQG-----H 545
Qy      638 EQAQSDVQKEANHLQSDGQQSVSKSEKHGCTKCFITQVESZSRHYR--AHSCYKR 695
Db      546 QSLSKJD-----CWAHARTHIPAEQLAKQCHCNFTYEKKHLEHYVRNHIGSPKPPQ 598
Qy      696 QCSFTAACQSLLHF--NTVCHEQODITTANGEDGHASTIKESPKCDFRVNLLTDC 753
Db      599 KAYNYCVTKMSLNSMKSHTWYFRMDCTYATKYCHS::KLHKKYN----- 646
Qy      754 SKYGPBPVSESVVREKJLEEKGDKLVKWTESSDDLRLNVTRGCADILRGSPSYQTASLG 813
Db      647 ---HRVPFGI-----EMSGGD-----SSPPTS DATIT 672
Qy      814 LPVGSTGEOTKTLDSPNVAAHLARP::YGLAVETKGFICGAPAGGKSGALPOQY 870
Db      673 FPLXKQCEIKTETVEPTVTSIAQPFPFNPMGNHGCLNFANHML----- 715
Qy      671 PASGENKSKDESQS-LRRRRSGVFCANCULTTKTSLWRKNANG-GVYVCAAGLYQM LHS 929
Db      716 -----NKHLDTGLMGLRSVM-PLKCSACDFVASSADEKWRHSMSHILNSSNPVTSIAS 970
Qy      930 PPLNIHKONGEQ---RRRTKRKNLPALQAELQINKOORGNSREQVNGSPLE-----RR 983
Db      771 YKSLND-----FSFSHVAPD-----NNDALESYCDQVKI 799

```

```

Cy 984 SCHLIESCHKA:FLFSGSKRISA QSSQKNSHQAQVFLVSGQLEHRRNQFHLGVINS 121
Db 900 DOUNITESH- ||||| CEEEDQGS- 817
Cy 1042 POESTGDPGNGSSVSEGGSGSGSPIEKYVRRPAK-HPNVSPPGSPIEKYQYFLGLPFV 110
Db 818 --SSAVPTGSSQ--SSGFEETKKCKSLSEGISARANGNSP-----M 859
Cy 1001 HNDFOSEADMLRFWSKYKLSVGPNGHYLSHVGP-PNFCQNVYVPYPTNLPPHPSAV- 115
Db 859 SNDSAVEKO-----GESADDAQPSPSDTTSVESP-----PLHSSIVAP 899
Cy 1157 ---GSDNDIPLDLAIKHSRPGP--TANGASKETKAPNVXNMGPLNVVTKVDRSTQDE 121
Db 899 PITPQPNFVQSILAQSLGSLFLANRPS----- 927
Cy 1213 LSTKCVHCG-VFLDSVYALHMYCHSGSPGPPQCS:COHLCCTDKYDFETH:CRGLHR 1268
Db 928 -APYCDCK:PDTCVQLDSHMR:PHTPGNPNCSQCOVQAFNELSLFALHMYQARHQ 962

RESULT 10
S05548
gap protein: hunchback - fruit fly (Drosophila virilis)
N:Alternate names: finger protein, hunchback
C:Species: Drosophila virilis
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998
C:Accession: S05548
R:Reiter, M.; Pfeifle, C.; Tautz, D.
EMBO J. 8, 1517-1525, 1989
A:Title: Comparison of the gap segmentation gene hunchback between Drosophila
A:Reference number: S05548; MUID:89356623; PMID:250459.
A:Accession: S05548
A:Molecule type: DNA
A:Residues: 1-516 <TRE>
A:Cross-references: EMBL:X15359; NID:g9178; PID:g9.79
C:Genetics:
A:Gene: hb
A:Cross-references: FlyBase:F9gn001316
C:Keywords: DNA binding; zinc finger

Query Match 1.2% Score 218.5 DB 2 Length 816

```



```
RESULT 12
157561
transcription factor GATA-4 - human
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Mar-1999
C:Accession: 157561; J04020
R:Yamagata, T.; Nishida, J.; Sakai, R.; Tanaka, H.; Hirano, K.; Mano, H.; Yazaki, M.; Cell. Biol. 15, 3830-3839, 1995
A:Title: Of the GATA-binding proteins, only GATA-4 selectively regulates the human interleukin-3 gene
A:Reference number: 157561; MUID:95311982; PMID:7791790
A:Accession: 157561
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-439 <RES>
A:Cross-references: GB:S78666; NID:G1042085; PID:G1042086
R:Huang, W.Y.; Cukerman, E.; Liew, C.C.
Gene 155, 219-223, 1995
A:Title: Identification of a GATA motif in the cardiac alpha-myosin heavy-chain-encoding gene
A:Reference number: JC4020; MUID:95217613; PMID:7721094
A:Accession: JC4020
A:Molecule type: mRNA
A:Residues: 'MAANHGPGRGAGGPPMHGAGASS', 'PVYLPTRVFSSVGLSLVLOGGAGSAGG', 64-68, 'P'
A:Cross-references: GB:L34357; NID:9508483
A:Experimental source: heart
C:Comment: This protein regulates a set of cardiac-specific genes and plays a crucial role in heart development
C:Genetics:
A:Gene: GDB:GATA4
A:Cross-references: GDB:433278; OMIM:600576
A:Map position: 8p23.1-8p23.1
C:Superfamily: transcription factor GATA-4; GATA-type zinc finger homology
C:Keywords: DNA binding; transcription factor; zinc finger
F:211-264/Domain: GATA-type zinc finger homology <GZF1>
F:214-238/Region: zinc finger GATA motif
F:265-318/Domain: GATA-type zinc finger homology <GZF2>
F:268-292/Region: zinc finger GATA motif

Query Match 3.1%; Score 215; DB 2; Length 439;
Best Local Similarity 33.0%; Pred. No. 5.6e-05;
Matches 69; Conservative 22; Mismatches 68; Indels 50; Gaps 9;

QY 873 SGENKSDSOSLLRRRGSGVFCANCLTTKTSIMRWKANGGVVNCNAGLYCKLHSTPEP 932
DB 245 NGINRPLIKPQRLSARRVGLSCANQTTTLWRNAGEPVCNAGLYKLRGVPR 304
QY 933 LNIQKNGEQIIRRTTRKLNPEALQAEQINLKQOR----GSNEEGVNGSPLEERSDHL 968
DB 305 LAMRKEG-----IQTRKRPKN-----LNKSKTPAAPSGSESLPPASGASSNSNAT 351
QY 989 TESHQREIPL---PSLSKYEAQSLTKSHSAQOPVLVSQTL-----DIHKMQPL 1035
DB 352 TSSSEEXRPITKTFGLSSH-----YGHSS-----VSQTFVSAMSGHGPHVLSAL 400
QY 1036 HIQIKSPQSTGDPGNSSSVSEKGSER 1064
DB 401 KL---SFQ-----GYASPVSQSPQTSK 420

RESULT 13
A40815
transcription factor GATA-2 (version 1) - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C:Accession: A40815
R:Lee, M.E.; Temizer, D.H.; Clifford, J.A.; Quattermost, T.
J. Biol. Chem. 266, 16188-16192, 1991
A:Title: Cloning of the GATA-binding protein that regulates endothelin-1 gene expression
A:Reference number: A40815; MUID:91340773; PMID:1714909
A:Accession: A40815
A:Molecule type: mRNA
A:Residues: 1-480 <LEE>
A:Cross-references: GB:W6889; NID:G192995; PID:AAA35868.1; PID:G182996
C:Genetics:
```

```
A:Gene: GDB:GATA2
A:Cross-references: GDB:133757; OMIM:137295
A:Map position: Xp11.23-Xp11.23
C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:292-345/Domain: GATA-type zinc finger homology <GZF1>
F:293-319/Region: zinc finger GATA motif
F:346-399/Domain: GATA-type zinc finger homology <GZF2>
F:349-373/Region: zinc finger GATA motif

Query Match 3.1%; Score 215; DB 2; Length 480;
Best Local Similarity 28.0%; Pred. No. 5.4e-05;
Matches 96; Conservative 26; Mismatches 89; Indels 106; Gaps 14;

QY 771 EEKDLKEKV-WTESSDDLRNVTRGADILRSPSYTQASLGLTPVSGTQCTKTLRD 829
DB 206 EDKGVKVCVSTESMKVE-----SGSP-----LRPGLAT-VGTQ----- 239
QY 830 SPNVEAAHLARPIYGLAVETKG-----FLOGAPAGEKSGALPQO----- 869
DB 240 ----PATHHPITPYESVYPAAHVSSGJFHPGGFLGGPASSFTPKORSKARSCSEGREC 295
QY 870 -----YPASGENKSKDESQSLRRRRRGSGVFCANCLTT 902
DB 296 VNCGATATP-WRDSGTSHY-CNACGLYHKMNGQNRPLIKRPLSAARRAGTCCANCTT 355
QY 503 KTSJMRKNANGGVVNCNAGLYCKLHSTPRPLNIKQNGEQIIRRTTRKLNPEALQAEQ 962
DB 356 TTKLRNANGDPVCNAGLYVYKLNVRPLTKKEG-----QTRNRKXSN-----K 403
QY 963 LNKQORSN-EE-----QVNGSPLEERS-----EDHJTE-SHQREIPLPSLSKYEAQ 1007
DB 404 SKSKGAGECFEELSKCKQKSEFFSAAA-AGHMAPVGHILPFPFSSHSHI-LPTPTPIHPS 462
QY 1008 GSLTKSH 1014
DB 453 SSLSFGH 469

RESULT 14
A56229
lymphoid transcription factor Ikaros/LyF-1, form IV - mouse
N:Alternate names: Ikaros/LyF-1 form I; Ikaros/LyF-1 form II; Ikaros/LyF-1 form III
C:Species: Mus musculus (house mouse)
C>Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 01-Dec-2000
C:Accession: A56229
R:Haem, K.; Ernst, P.; Lo, K.; Kim, G.S.; Turck, C.; Smale, S.T.
Mol. Cell. Biol. 14, 7111-7123, 1994
A:Title: The lymphoid transcription factor LyF-1 is encoded by specific, alternative-y
A:Reference number: A56229; MUID:95021239; PMID:7935426
A:Accession: A56229
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-392 <HAH>
A:Cross-references: GB:S74517
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: alternative splicing

Query Match 3.1%; Score 213; DB 2; Length 392;
Best Local Similarity 21.4%; Pred. No. 6.3e-05;
Matches 122; Conservative 65; Mismatches 169; Indels 214; Gaps 26;

QY 728 EDGHAISTI--KEEPKIDFRVYNLLTPDSKMGEPVSVVKKREKLEKGLKEKWTESS 785
DB 4 DEGDMSQVSGKSEPPVSC-----TPDEG-----DEPMVPVEDLSTTSGAQQ----NSK 48
QY 786 SDDLNRNVTRGAD-----IL--RGSPSYTQASLGLLTPVSGTCEOTK-LRDSPNVEAA 836
DB 49 SD--RGVAVYAGDGRDFHAIISDRGMS-----NVKVTCSDENGRAC 91
QY 837 HLARPIYGLAVETKGF-QGAPAGKEKSGALPQQVYPASGENKS---KQESQSLRRRRRGSG 893
DB 92 EM-----NGECAEDLRMLDASGKMGSHRQGGSAJ-----SG 126
```

```
QY 894 VFCANCLTTKTSLRKNANGGVYCNACGLYQKHLHSTPRPLN::KQNGEQIIRRRTRKEL 953
Db 127 VGGI-----R.PNGKGLXCDICGIV-----CI 147
QY 954 NPEALQAEQLNKQORGSNEEQVNGSP::ERRSEDLHTESHORE::PLPSLSKYEACGSLTKS 1013
Db 148 GPNVLM-----VHKRSH---TKKCLSDMPYS-ANYEKEDNMT-S 182
QY 1014 HSAQQPVLSQTLTDIHKRMOP::HIQIKSPOESTGDFGNSSSVSEGGKSGSERGSP::EKYMR 1073
Db 183 HVMQAINNAINVYLGAEPL---VQTP-----FGSEVVVSVSSMYC 223
QY 1074 PAKHNYSPGSP::EKYQVYPLGLPFLVHNDPQSEADWLRFWFKYKLSVPCNPHYLSHVFG 1133
Db 224 LHKPPSDGPPRS-----MHSQAQDAVENLLLSKAK-SVS-----SREFA 261
QY 1134 LP-NPCQNVYP-----YPTFNLPHPFSAVGSNDIPLDLAIKHSRPGTKAG 1179
Db 262 SPNSQCDSTDTESNAEQRSGLYLTNHNPH-----ARNG 298
QY 1180 ASKEKTKAPENVKNEGLNVVYKTEKVDNSTQDELST-KCVHCGIVFLDEVVYALHVSCHG 1238
Db 299 LALKEQRAVEVLRAAENSQDAFRVSVTSGEQLKVYKCEHCRVLFELCHVWYTIHMGCHG 358
QY 1239 DSGPPQCSICQHLCTDKYDTTHIQRLHR 1268
Db 359 FRDPFECMCGYHSQDRYEFSSHITRGEHR 388

RESULT 15
S40382
box A-binding factor - fruit fly (Drosophila melanogaster)
N;Alternate names: ABF; transcription factor dGATAB
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 06-Sep-1996 #text_change 16-Jul-1999
C;Accession: S40382
R;Abel, T.; Michelson, A. M.; Maniatis, T.
Development 119, 623-633, 1993
A;Title: A Drosophila GATA family member that binds to Adh regulatory sequences is expressed in the developing eye
A;Reference number: S40382; MUID:94244465; PMID:8187633
A;Accession: S40382
A;Molecule type: mRNA
A;Residues: 1-779 <ABE>
A;Cross-references: EMBL:X76217; NID:9441491; PIDN:CAA53607.1; P/D:9441492
C;Comment: This transcriptional activator is the earliest known marker of the developing eye
C;Genetics:
A;Gene: FlyBase:srp
A;Cross-references: FlyBase:FBgn0003507
C;Superfamily: box A-binding factor; GATA-type zinc finger homology
C;Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc finger
F;316-369/Domain: GATA-type zinc finger homology <GZF>
F;319-343/Region: zinc finger GATA motif

Query Match 3.1%; Score 213; DB 1; Length 779;
Best Local Similarity 25.3%; Pred. No. 0.00016;
Matches 80; Conservative 38; Mismatches 98; Indels 104; Gaps 12;

QY 849 TKGFLQAGAPAGGKSGALPOQYPASG-----ENKSKDESQSLRRRRRGSGVFCAN 898
Db 263 TSGAASSYSCPGSNATSAATSAVASGTAAATATLDEHVSRRRRLSASKR-AGLSCSN 321
QY 899 CLTTKTSLRKNANGGVYCNACGLYQKHLHSTPRPLN::KQNGEQIIRRRTRKELNPEAL 958
Db 322 CHTTHTSLWRNPAGEPVCNACGLYKLSVPRLTMKKDT-----IQKSKRK---PKGT 373
QY 959 QAEQ-LNKQQRGSNEEQVNGSFLER-RSEDLHTESHOREI-----P 997
Db 374 KSEKSKGSKUALNAIMESGSLVTNCHNVGVLDSSQMDVNDOMKPOLDLKPNYSYSSQP 433
QY 998 LPSLSKYEAQGSLS-----TKSHSAQQPVLVS--- 1023
Db 434 QQLPQYQQQQQQLVADNADHSSAASPMSMGSTSLSPSAMSHQHQTHPHQQQQQQLCGSWT 493
```

```
QY 1024 ----QTLQ-HKR-----MOPLHIOIKSP--QESTGDPGNKSSSVSECKG 1060
Db 494 CRPTQTTXCRBRSTGSS-SSSRRAACSTHFAHPJLHQHPSPTHOLHNNNNKSSLLFNNN 553
QY 1061 SSERGSP-----EKYMR 1073
Db 554 NNNSSNNNNNNKLLQKYLQ 573

Search completed: October 29, 2003, 11:44:08
Job time : 76 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:56 ; Search time 39 Seconds
(without alignments)

1544.646 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MTRKKVPLPNVASEGGQI.....IORGJHRNNAQVEKNGKPK 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6843	99.9	1281	1 TRPS_HUMAN	Q9ufh7 homo sapien
2	6424	93.8	1281	1 TRPS_MOUSE	Q925h1 mus musculu
3	5107	74.5	1271	1 TRPS_XENLA	Q902s6 xenopus lae
4	273.5	4.0	757	1 HUNB_DROME	O62538 drosophila
5	271	4.0	758	1 HUNB_DROME	P05084 drosophila
6	269	3.9	519	1 IKAR_HUMAN	Q13422 homo sapien
7	266.5	3.9	767	1 HUNB_DROOR	O62537 drosophila
8	251.5	3.7	1829	1 Y296_HUMAN	O15015 homo sapien
9	249.5	3.6	759	1 HUNB_DROYA	O62541 drosophila
10	248.5	3.6	1114	1 TCF8_CHICK	P26197 gallus gall
11	245	3.6	3703	1 ABFI_HUMAN	Q15911 homo sapien
12	241	3.5	452	1 GAT2_XENLA	P23770 xenopus lae
13	238	3.5	1350	1 XFIN_XENLA	P08045 xenopus lae
14	234.5	3.4	971	1 AREA_GIBFU	P78688 gibberella
15	234	3.4	517	1 IKAR_MOUSE	Q03267 mus musculu
16	234	3.4	782	1 ZFY1_MOUSE	P10925 mus musculu
17	230	3.4	518	1 IKAR_CHICK	Q42410 gallus gall
18	223.5	3.3	1687	1 Z142_HUMAN	P52746 homo sapien
19	222.5	3.2	466	1 GAT2_CHICK	P23824 gallus gall
20	222.5	3.2	982	1 HBL1_CAMEL	Q9xyd3 caenorhabdi
21	221.5	3.2	817	1 HUNB_MOUSE	Q01778 musa domes
22	220.5	3.2	3726	1 ABFI_MOUSE	Q61329 mus musculu
23	218.5	3.2	816	1 HUNB_DROVI	P13361 drosophila
24	218	3.2	522	1 IKAR_ONCMY	O13089 oncorhynchu
25	218	3.2	783	1 ZFY2_MOUSE	P20662 mus musculu
26	217	3.2	397	1 GAT5_HUMAN	Q9bwx5 homo sapien
27	215	3.1	442	1 GAT4_HUMAN	P43694 homo sapien
28	213	3.1	480	1 GAT2_HUMAN	P23769 homo sapien
29	213	3.1	779	1 SRP_DROME	P52172 drosophila
30	212.5	3.1	391	1 GAT6_XENLA	P70005 xenopus lae
31	212.5	3.1	440	1 GAT4_RAT	P46152 rattus norv
32	212.5	3.1	526	1 HELI_HUMAN	Q9uk57 homo sapien
33	212.5	3.1	526	1 HELI_MOUSE	P91183 mus musculu

34 212 3.1 404 1 GAT5_MOUSE
35 212 3.1 480 1 GAT2_MOUSE
36 212 3.1 1507 1 PRDF_HUMAN
37 209.5 3.1 392 1 GAT4_XENLA
38 209.5 3.1 392 1 GAT4_XENLA
39 208 3.0 436 1 GAT3_BRARE
40 208 3.0 1342 1 Z135_HUMAN
41 207.5 3.0 444 1 GAT3_CHICK
42 207 3.0 443 1 GAT3_HUMAN
43 207 3.0 443 1 GAT3_MOUSE
44 207 3.0 1109 1 TCF8_RAT
45 207 3.0 1124 1 TCF8_HUMAN

ALIGNMENTS

RESULT 1
ID TRPS_HUMAN STANDARD; PRT: 1281 AA.
AC Q9UHF7; Q9UHF7; Q9UHF7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger transcription factor Trps. Zinc finger protein GC73;
DE (Tricho-rhino-phalangeal syndrome type I protein).
GN TRPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX MCB; TaxID=9606;
IL [1]_TaxID=9606;
RN SEQUENCE FROM N.A.
RX MEDLINE=2008282; PubMed=1061511;
RA Momeni P., Gloeckner G., Schmidt O., von Holtum D., Albrecht B.,
RA Gillissen-Kaesbach G., Hennekam R.C.M., Meinecke P., Zabel B.,
RA Rosenthal A., Hortschmke B., Lueddecke H.-J.;
RT "Mutations in a new gene, encoding a zinc-finger protein, cause
tricho-rhino-phalangeal syndrome type I";
RL Nat. Genet. 24:71-74(2000).
RN [2]
RX SEQUENCE FROM N.A.
RA Chang G.T.C., Steenbeek M., Schippers E., Blok L.C., van Weerden W.M.,
RA van Aalwijk D.C.G., Eussen S.H.C., van Steenbrugge G.J.,
RA Brinkman A.O.;
RT "Characterization of a zinc-finger protein and its association with
apoptosis in prostate cancer cells";
RL J. Natl. Cancer Inst. 92:1414-1421(2000).
RN [3]
RX SEQUENCE OF 495-1279 FROM N.A.
RA Isocai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Toshiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Makamura Y., Nagahata K., Masuho Y., Oshima A.;
RT "NEO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RX VARIANTS TRPS3 ASF-934; PRO-901; GLN-908; PRO-909 AND THR-919.
RA MEDLINE=21063177; PubMed=1112658;
RA Lueddecke H.-J., Schaper J., Meinecke P., Momeni P., Gross S.,
RA von Holtum D., Hirsch H., Abramowicz M.J., Albrecht B., Forderer A.,
RA Christen H.-J., Clauser U., Devriendt M.J., Pastnacht E., Forderer A.,
RA Friedrich U., Godtschap T.H.C., Greive X., Hamm H., Hennekam R.C.M.,
RA Rinkel G.K., Hoeftzenbein M., Kayserli H., Matjewska P., Mathieu M.,
RA McLeod R., Miro A.T., Moog J., Nagai T., Nishikawa K., Oerstavik K.H.,
RA Pleckch E., Saitz C., Schmidtke J., Traubaeerg L., Tsukahara M.,
RA Wittwer B., Zabel B., Gillesen-Kaesbach G., Horsthenke B.;
RT "Genotypic and phenotypic spectrum in tricho-rhino-phalangeal syndrome
types 1 and 11";

Am. J. Hum. Genet. 68:81-91(2001).

[5] VARIANT TRPS3 GLN-908.

MEDJINE=21665516; PubMed=11807863;

Kobayashi H., Hino M., Shimodaira M., Iwakura T., Ishihara T.,

RA Ikebuchi K., Ozawa Y., Nakao K., Kurahachi H.,

RA "Missense mutation of TRPS1 in a family of tricho-rhino-phalangeal

PT syndrome type III.";

RT Am. J. Med. Genet. 107:26-29(2002).

CC -!- FUNCTION: Transcriptional repressor. May act to restrict

CC expression of GATA-regulated genes at selected sites and stages in

CC vertebrate development. Might be involved in prostate cancer

CC apoptosis.

CC -!- SUBUNIT: Binds specifically to GATA sequences (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in the adult. Found in

CC fetal brain, lung, kidney, liver, spleen and thymus. More highly

CC expressed in androgen-dependent than in androgen-independent

CC prostate cancer cells.

CC -!- DISEASE: Defects in TRPS1 are the cause of tricho-rhino-phalangeal

CC syndrome type I (TRPS1) [MIM:190350], an autosomal dominant

CC disorder characterized by craniofacial and skeletal abnormalities.

CC It is allelic with tricho-rhino-phalangeal type III. Typical

CC features include sparse scalp hair, a bulbous tip of the nose,

CC protruding ears, a long first phalanx and a thin upper vermillion

CC border. Skeletal defects include cone-shaped epiphyses at the

CC phalanges, hip malformations and short stature.

CC -!- DISEASE: Defects in TRPS2 are a cause of tricho-rhino-phalangeal

CC syndrome type II (TRPS2) [MIM:150230]. TRPS2 is a contiguous gene

CC syndrome due to deletions in chromosome 8q24.1 and resulting in

CC the loss of functional copies of TRPS1 and EXT1.

CC -!- DISEASE: Defects in TRPS3 are the cause of tricho-rhino-phalangeal

CC syndrome type III (TRPS3) [MIM:190351], an autosomal dominant

CC disorder characterized by craniofacial and skeletal abnormalities.

CC It is allelic with tricho-rhino-phalangeal type I. In TRPS3 a more

CC severe brachydactyly and growth retardation are observed.

CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.

CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

EMBL; AF183810; AAF23614.1; -

DR EMBL; AF178030; -; NOT ANNOTATED_CDS.

DR EMBL; AF264784; AAG21134.1; -

DR EMBL; AK000948; BAA91441.1; -

DR HSSP; P17679; IGNF.

DR Genew; HGNC:12340; TRPS1.

DR MIM; 604386; -

DR MIM; 190350; -

DR MIM; 190351; -

DR MIM; 150230; -

DR InterPro; IPR007087; Znf_C2H2.

DR InterPro; IPR000679; Znf_GATA.

DR Pfam; PF00320; GATA.1.

DR Pfam; PF00096; zf-C2H2.4.

DR PRINTS; PR00619; GATAZNFINGER.

DR SMART; SM00355; Znf_C2H2; 9.

DR SMART; SM00401; Znf_GATA; 1.

DR PROSITE; PS00344; GATA_2N_FINGER_1; 1.

DR PROSITE; PS00114; GATA_2N_FINGER_2; 1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.

DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.

CC Transcription regulation; Repressor; Zinc-finger; DNA-binding;

KW Nuclear protein; Metal-binding; Repeat; Disease mutation.

KW 2N_FING 222 247 C2H2-TYPE 1 (ATYPICAL).

FT 2N_FING 333 358 C2H2-TYPE 2 (ATYPICAL).

FT 2N_FING 614 637 C2H2-TYPE 3 (ATYPICAL).

FT	2N_FING	666	689		C2H2-TYPE 4.
FT	2N_FING	692	715		C2H2-TYPE 5.
FT	2N_FING	896	920		GATA-TYPE.
FT	2N_FING	1215	1237		C2H2-TYPE 6.
FT	2N_FING	1243	1267		C2H2-TYPE 7.
FT	VARIANT	894	894		V->D (in TRPS3; in heterozygous status
FT					has a milder effect causing TRPS1).
FT					/Ffid=VAR_012807.
FT	VARIANT	901	901		T->P (in TRPS3; severe).
FT					/Ffid=VAR_012808.
FT	VARIANT	908	908		R->P (in TRPS3; severe).
FT					/Ffid=VAR_012809.
FT	VARIANT	908	908		R->Q (in TRPS3).
FT					/Ffid=VAR_012810.
FT	VARIANT	919	919		A->T (in TRPS3).
FT					/Ffid=VAR_012811.
FT	CONFLICT	115	115		S->F (in REF. 1; AAF23614).
FT	CONFLICT	1276	1279		NGXP->KHKT (in REF. 3).
FT	SEQUENCE	1281	AA: 141520	MM: 215780	4F55EB71CC CRCE4;

Query Match 99.9%; Score 6843; DB 1; Length 1281;

Best Loca. Similarity 99.9%; Pred. No. 0;

Matches 1260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MVRKNPFLRNVASEGQILEPIGTSKVSQGNKKEFSACQXSENTQSCAAELNHKEEH	60
DB	1	MVRKNPFLRNVASEGQILEPIGTSKVSQGNKKEFSACQXSENTQSCAAELNHKEEH	60
QY	61	SLHVCCPSSSSKDLKSAVLSEKAGFNYESPSGGRFESPHDEVTDENMLASFPAAGG	120
DB	61	SLHVCCPSSSSKDLKSAVLSEKAGFNYESPSGGRFESPHDEVTDENMLASFPAAGG	120
QY	121	VCEPLKSPQRAEDDPOQMACTPSGDSLETKEQKMSFKATEETGQAQSQQAQCGLSPV	180
DB	121	VCEPLKSPQRAEDDPOQMACTPSGDSLETKEQKMSFKATEETGQAQSQQAQCGLSPV	180
QY	181	SVASKNPQVPSDQGVRLNKSKTLLVNDNPDPAPLSPQLQCFKNCIGYGYGNDPTDLI	240
DB	181	SVASKNPQVPSDQGVRLNKSKTLLVNDNPDPAPLSPQLQCFKNCIGYGYGNDPTDLI	240
QY	241	KHFRKYLHGLHNRTRCDAELDSKTLALHNVCVSHSKFCQKVARSVESVGLQGINSRPV	300
DB	241	KHFRKYLHGLHNRTRCDAELDSKTLALHNVCVSHSKFCQKVARSVESVGLQGINSRPV	300
QY	301	LLNGTYDVQVTSSTGTFIGIGRKTPTDCGNTKYFRCKFCNTFTMGNSSTELSQHFLQHPN	360
DB	301	LLNGTYDVQVTSSTGTFIGIGRKTPTDCGNTKYFRCKFCNTFTMGNSSTELSQHFLQHPN	360
QY	361	KIKASLPSSEVAKPSEKNSKSIIPALQSSDSDGLGKWQDKITVAGDCTPVGYSVFKPL	420
DB	361	KIKASLPSSEVAKPSEKNSKSIIPALQSSDSDGLGKWQDKITVAGDCTPVGYSVFKPL	420
QY	421	DSSRCNGTEATSYVCKPFCFSCESSSSSLKLLHYGHQHGAVQSGGNPELNDKLSRGSV	480
DB	421	DSSRCNGTEATSYVCKPFCFSCESSSSSLKLLHYGHQHGAVQSGGNPELNDKLSRGSV	480
QY	481	INQNDLAKSSEGETMTKTDKSSGAKKXDFSSKGAEDNMVTSYNCQCFDPRYSKSHGPDV	540
DB	481	INQNDLAKSSEGETMTKTDKSSGAKKXDFSSKGAEDNMVTSYNCQCFDPRYSKSHGPDV	540
QY	541	IVVGPLLRHYQQLHNIHKTIKCPFCPRGLCSPEKHLGEITYPFACRKNSCHCALLLL	600
DB	541	IVVGPLLRHYQQLHNIHKTIKCPFCPRGLCSPEKHLGEITYPFACRKNSCHCALLLL	600
QY	601	HLSPGAGSSRVKHQCHQCSFTTDPDVVLLFHVSVHESQASDVKQEAHLQSGSDGQSV	660
DB	601	HLSPGAGSSRVKHQCHQCSFTTDPDVVLLFHVSVHESQASDVKQEAHLQSGSDGQSV	660
QY	661	KESHEHSCTKCDFITQVEEESISRHYRAHSCYKRCQCSFTAAQTQSLLEHFNTHVCOEQD	720
DB	661	KESHEHSCTKCDFITQVEEESISRHYRAHSCYKRCQCSFTAAQTQSLLEHFNTHVCOEQD	720
QY	721	TTANGSEGDGHA-STIKKEEPKIDFRVYNLLTPDSKMGEPVSESVYKSEKLEKQGLKEKV	780

```

Db 721 ITTANGEEDGHALST:KEPKIDFRVYNLLTPSKVGEPEVSVEKLEEKDGLKEKV 780
QY 781 WTSSDDLRNVTWRGADILRSPSVTQASLGLLTPVSGTQOTKTJROSPNVEAAHLAR 840
Db 781 WTSSDDLRNVTWRGADILRSPSVTQASLGLLTPVSGTQOTKTJROSPNVEAAHLAR 840
QY 841 PLYGLAVENKGFLOGAPAGEKSGALPOQYPASGENKSKDESQSLRRRSGSVFCANCL 900
Db 841 PLYGLAVENKGFLOGAPAGEKSGALPOQYPASGENKSKDESQSLRRRSGSVFCANCL 900
QY 901 TTKTSLWRKNANGGYVCNACGLYQKHLSTPRPLNIIKONNGEOTIRRTKRLNPEALCA 960
Db 901 TTKTSLWRKNANGGYVCNACGLYQKHLSTPRPLNIIKONNGEOTIRRTKRLNPEALCA 960
QY 961 EQLNKOORSGNEBQVNGSPLRRSEDLTESHOREIPJPSLSKYEAQGSILTKSHSAQPV 1020
Db 961 EQLNKOORSGNEBQVNGSPLRRSEDLTESHOREIPJPSLSKYEAQGSILTKSHSAQPV 1020
QY 1021 LVSOTLDIHKWOPLHIQIKSPQESTGDRQNSSSVSEGGSSERGSPIEKYWRPAKHPNY 1080
Db 1021 LVSOTLDIHKWOPLHIQIKSPQESTGDRQNSSSVSEGGSSERGSPIEKYWRPAKHPNY 1080
QY 1081 SPGSPTEIKYQYPLFGLPFFVHNDPQSEADWLRFWSKYKLSVPGNPHYLSHVPLNPPCN 1140
Db 1081 SPGSPTEIKYQYPLFGLPFFVHNDPQSEADWLRFWSKYKLSVPGNPHYLSHVPLNPPCN 1140
QY 1141 YVPYPTFNLPFHFSVAGSNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNKGPJNVY 1200
Db 1141 YVPYPTFNLPFHFSVAGSNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNKGPJNVY 1200
QY 1201 KTEKVDSTRQDELSTKCVHCGIVFLDEVYALHMSCHGSDGPGQCSICQHLCTDKYDFT 1260
Db 1201 KTEKVDSTRQDELSTKCVHCGIVFLDEVYALHMSCHGSDGPGQCSICQHLCTDKYDFT 1260
QY 1261 HQRGLHRNNAQVEKNGKPKPE 1281
Db 1261 HQRGLHRNNAQVEKNGKPKPE 1281

```

RESULT 2

```

TRPS_MOUSE
ID TRPS_MOUSE STANDARD; PRT; 1281 AA.
AC Q925H1.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger transcription factor Trps1.
GN TRPS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Embryonic gut;
RX MEDLINE=21181829; PubMed=11285235;
RA Malik T.H., Shoichet S.A., Latham P., Kroll T.G., Peters L.L.,
RA Shivdasani R.A.;
RT "Transcriptional repression and developmental functions of the
RT atypical vertebrate GATA protein Trps1.";
RL EMBO J. 20:1715-1725(2001).
CC -!- FUNCTION: transcriptional repressor. May act to restrict
CC expression of GATA-regulated genes at selected sites and stages in
CC vertebrate development.
CC -!- SUBUNIT: Binds specifically to GATA sequences.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: In the embryo, expression is detected in both
CC visceral and skeletal tissues. Found in the maxilla, mandible,
CC snout, prospective phalanges and in the femoral head within the
CC developing hip. Also expressed in the hair follicles.
CC -!- DEVELOPMENTAL STAGE: Detected prior to E7.5, with peak levels at
CC around E11.5. In the developing limbs and face, levels are highest

```

```

CC at E13.5 and decline dramatically thereafter.
CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an e-mail to license@ebi.ac.uk).
CC
CC EMBL: AF346936; AAK39528.1;
CC MGD: MGI:192616; Trps1;
CC InterPro: IPR007087; Znf_C2H2.
CC Pfam: PF00320; GATA_1;
CC Pfam: PF00396; Zf_C2H2_5;
CC PRINTS: PR00619; GATAZNFINGER.
CC SMART: SM00355; Znf_C2H2_9.
CC SMART: SM00451; Znf_GATA_1;
CC PROSITE: PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE: PS00114; GATA_ZN_FINGER_2; 1.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
CC PROSITE: PS01571; ZINC_FINGER_C2H2_2; 2.
CC Transcription regulation; Repressor; Zinc-finger; DNA-binding;
CC Nuclear protein; Metal-binding; Repeat.
CC ZN_FING 222 247 C2H2-TYPE 1 (ATYPICAL).
CC FT ZN_FING 333 358 C2H2-TYPE 2 (ATYPICAL).
CC FT ZN_FING 614 637 C2H2-TYPE 3 (ATYPICAL).
CC FT ZN_FING 666 689 C2H2-TYPE 4.
CC FT ZN_FING 692 715 C2H2-TYPE 5.
CC FT ZN_FING 996 920 GATA-TYPE 6.
CC FT ZN_FING 1215 1237 C2H2-TYPE 7.
CC FT ZN_FING 1243 1267 C2H2-TYPE 7.
CC SEQUENCE 1281 AA; 141034 MW; FCA7190C2357898 CRC64;

```

Query Match: 93.8%; Score 6424; DB 1; Length 1281;

Best Local Similarity 93.1%; Pred. No. 0;

Matches 1192; Conservative 39; Mismatches 51; Indels 0; Gaps 0;

```

QY 1 MYRKXNPLRNVASEGEGCILEP-GTSSKVSQKKEFSADOMSENTDQSDAAELNHKEE 60
DB 1 MYRKXNPLRNVASEGEGCILEP-GTSSKVSQKKEFSADOMSENTDQSDAAELNHKEE 60
QY 61 SLHVQDPSSSSKKOLKSAVLSEKAGFNYESKGNFPSPFPHDEVTDNMJAFSPAAAG 120
DB 61 SLHVQDPSSSSKKOLKSAVLSEKAGFNYESKGNFPSPFPHDEVTDNMJAFSPAAAG 120
QY 121 VCEPLKSPQRAEADDPQDMACTPSGDSLETKEEDOKYSPKATEETGQAGSGGACQGLSPV 180
DB 121 VCEPLKSPQRAEADDPQDMACTPSGDSLETKEEFHMGFKATEETGPVQSGGACQGLSPV 180
QY 181 SVASKNPQVPSDGGVRLNKSXTDLLVNDNPAPALSPPELQDFKCNICGYGYGNDPTDLI 240
DB 181 SVASKNPQVPSDGGVRLNKSXTDLLVNDNPAPALSPPELQDFKCNICGYGYGNDPTDLI 240
QY 241 KHFKYHLGLHNRTRQCAELDSKILALHNNYQVFSHKSDFQKYNRSVFSGLVQDINSRPF 300
DB 241 KHFKYHLGLHNRTRQCAELDSKILALHNNYQVFSHKSDFQKYNRSVFSGLVQDINSRPF 300
QY 301 LLNGTYDVQVTSGGTF-GIGRKTDCQGNKYFCKFCNFYMGNSSTELFCHPLOTHPN 360
DB 301 LLNGTYDVQVTSGGTF-GIGRKTDCQGNKYFCKFCNFYMGNSSTELFCHPLOTHPN 360
QY 361 KIKASLPSEVAKPKSEKSNKSIIPALOSSGSDGKQODKTVKAGDDTPVGYVPIKPL 420
DB 361 KIKASLPSEVAKPKSEKSNKSIIPALOSSGSDGKQODKTVKAGDDTPVGYVPIKPL 420
QY 421 DSSQNGTEATSYWCKPFCSPSCSSSLKLEHYGKHQGVQSGGKNPENLNDKLSRGSV 480
DB 421 DSSQNGTEATSYWCKPFCSPSCSSSLKLEHYGKHQGVQSGGKNPENLNDKLSRGSV 480

```



```
Qy 578 LGBITVFAKCKNCCHCALLLHLSPGAAGSRVKGHCQCFTTPDVLPHFVSVH 637
Db 146 LRPSPOTPTSTASTIAPVAV-----ATGSS-----EKLQALTPMDVT--PKPSA 199
Qy 638 ESQASDVQKQANHLQSGDGOOSVK---ESKEH----- 666
Db 190 KSSQSNIEPEKHDQMSNSEDKMYAESDEDDTNRMPYNSGKVKMKYKCTGQVVA 249
Qy 667 -----SCTKCDFTQVEEISRHVR--RAHSCVKKQCQCSFTAAATQ 705
Db 250 TKVDFWATHTMKPKDILQCPKCPFTFEPKHLEHVRKHQKQPCQDCSYTCVWKS 329
Qy 706 SLLEHFN-----VHCQSQOIT-----ANGEEDGHAISTIKEB-----PKIFRVY 747
Db 310 MLKSHRSHSVYQYRCACDQYATKYCHSPKFLHRKYGRKPGVMVDEDGTPEFNSLVIDV 369
Qy 748 NLLT-PSKMGEPVSESVVREKLEEDGGLKEKVTWTESSDCLLVNTRGADILAGSPSY 826
Db 370 GTRGPKSKNGGFIASG----- 866
Qy 807 TOASLGLTTPVSGTQETKLRDPSNPNVAAHLARPIYGLAVETKGFQAGAGGEXSGAL 866
Db 392 SRKS-----NVAAPVAPQOQSQAPQVATSHLSAALQGFPL-----VQSNAPPAASPVJ 441
Qy 867 PQQYPASGENK--SKDESQSLRRRSGGVFCANCLTTSKLRKXANGGVYCNACGLYQ 924
Db 442 P--LPASPAKSVASVEQTPLSPG-----ANLLPPLASLLQQRNMMAFF-----PYW 486
Qy 925 KLHSTPRPLNIQNNGEQIIRTRKRLNPEALQAEQLNKGQSGNEEQVNGSPLEERS 984
Db 487 NLM-----LQMLAQQAQAVL-----ACLSPRVREQLQQNQOQSDNEEDQDEVERKS 536
Qy 985 EDHLTESHOREIPLPSLSKYEAQGSLSKSAQPVLSQTLQHKMQLPHLQIKSPQE 1044
Db 537 VDSAMD-----SQGTPVKEDBQQQ-----QQQFLAXNLKVEEB 571
Qy 1045 STGDPGNSVSEBKGSSRGSPTEKYMRFAKHFNYSPPGSPTEKYQYPLFGVFNHDF 1104
Db 572 AT-PLMSSNASRKG---RVKLDTLLQ----- 597
Qy 1105 QSEADMLRFKSKYKLSVPGNPHYLHVPLNPQNVVPTNLPHFSAVGSNDZPL 1164
Db 598 RSEA-----MTSPBOLKVPSTP-----MPTASSPIAGRKMP-----EBHCSGTSSADE--- 644
Qy 1165 DLAIKH-----SRPGPTANGASKETKAPPNVKNVKEPLNVVYKTEKYDRSTQDE 1212
Db 642 SMERPHVRQANTSASTSSGNSNASSNNGSSNNGSSNGTTSAAVAPPSGTTPAAG 701
Qy 1213 LSTKCVHGVFLDEVYALHMSCHGSGFPQCSICQHLCTDKYDFTTHIORGLH 1267
Db 702 AYECKYCDIFFKDAVLVYTHMGYHSCDDVFCKNMCCKCDGPGVGLFVHMARNAH 756
RESULT 5
HUNB_DROME STANDARD; PRT; 758 AA.
AC P05084; Q24018;
DT 01-APR-1988 (Rel. 07, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hunchback protein.
GN HB CR CG9786.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RA Tautz D., Lehmann R., Schuerch H., Schuh R., Seifert E.,
RA Kienlin A., Jones K., Jaekle H.;
```

```
RT "Finger protein of novel structure encoded by hunchback, a second
RL member of the gap class of Drosophila segmentation genes.";
RN Nature 327:383-389(1987).
RP REVISION TC 525.
RA Tautz D.
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=94259232; PubMed=8203478;
RA Margolis J.S., Borowsky V., Shim C.W., Posakony J.K.;
RL "A small region surrounding the distal promoter of the hunchback gene
RT directs maternal expression.";
RN Dev. Biol. 163:381-388(1994).
RP SEQUENCE FROM K.A.
RC STRAIN=Berkely;
RX MEDLINE=20396506; PubMed=10731132;
RA Adams M.D., Geiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaral-Andes P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vardell M.D., Zhang C., Chen L.X.,
RA Brandon S.C., Rogers J.H.C., Blazer R.G., Champe V., Pfeiffer B.D.,
RA Abril J.F., Adaya A., An H.-J., Andrews-Flannock C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale C., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bermet B.P., Bhandari D., Bolshakov S.,
RA Brkova D., Botchan N.R., Bouck J., Brockstein P., Brettler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry C.M., Cawley S., Dahlke C., Davenport J.B., Davies P.,
RA De Fabris B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.X.,
RA Dedson K.J., Evangelista C.C., Petraz C., Ferreira S., Fleischmann W.,
RA Foster C., Garg A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goddek A., Gong P., Gorrell J.H., Gu Z., Guar P., Harris M.,
RA Harris N.L., Harvey D., Heiman T., Hernandez J.S., Houck J.,
RA Houston D., Houston K.A., Howard T., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.E., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei J., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maize J., McIntosh T.C., McLeod X.P., McPherson J.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris C., Moshrefi A.,
RA Mount S.X., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzo D.M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Sider-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Kassam D.A., Weistock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu J., Yang S., Yao Q.A.,
RA Ye C., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2003).
RP SEQUENCE OF 281-349 FROM N.A.
RX MEDLINE=93066327; PubMed=1438276;
RA Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.;
RL "Evolutionary conservation pattern of zinc-finger domains of
RT Drosophila segmentation genes.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786(1992).
RP POLYMERASE.
RA Tautz D., Nigro J.;
RT "Microevolutionary divergence pattern of the segmentation gene
RL hunchback in Drosophila.";
RN Mol. Biol. Evol. 15:1403-1411(1998).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: IN EMBRYO. EXPRESSION OF MATERNAL TRANSCRIPT
```


RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein X.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.K.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J.J., Helton E., Kettenan M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting X., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Brakesley R.W., Touchman J.W., Schuetz J., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Green J., Myers R.M.,
RA Butterfield V.S.N., Krywinski M., Skalska U., Smalus D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.,
RA human and mouse cDNA sequences.";
RT "Generation and initial analysis of more than 15,000 full-length
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC THE CD3-Delta GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC CONTROL ELEMENT IN THE TBT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC PROMOTOR AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED
CC DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=Ik1;
CC IsoId=Q13422-1; Sequence=Displayed;
CC Name=Ik2;
CC IsoId=Q13422-2; Sequence=VSP_006848;
CC Name=Ik3;
CC IsoId=Q13422-3; Sequence=VSP_006850;
CC Name=Ik4;
CC IsoId=Q13422-4; Sequence=VSP_006847, VSP_006850;
CC Name=Ik5;
CC IsoId=Q13422-5; Sequence=VSP_006852;
CC Name=Ik6;
CC IsoId=Q13422-6; Sequence=VSP_006849;
CC Name=Ik7;
CC IsoId=Q13422-7; Sequence=VSP_006851;
CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THYMUS, SPLEEN AND
CC PERIPHERAL BLOOD LEUKOCYTES AND LYMPH NODES. LOWER EXPRESSION IN
CC BONE MARROW AND SMALL INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/Ikaros1258.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U40462; AAC50459.1; -;
DR EMBL: S80876; AAB50693.1; -;
DR EMBL: BC018349; AAB58349.1; -;
DR HSP: F15822; IBB0.
DR TRANSFAC: T02702; -;
DR Genew: HGNC:13176; ZNFN1A1.
DR MIM: 603023; -;
DR GO: GO:0003677; P:DNA binding activity; TAS.
DR GO: GO:0007498; P:mesoderm development; TAS.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 6.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0057; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT ZN_FING 117 139 C2H2-TYPE 1.
FT ZN_FING 145 167 C2H2-TYPE 2.
FT ZN_FING 173 195 C2H2-TYPE 3.
FT ZN_FING 201 224 C2H2-TYPE 4.
FT ZN_FING 462 484 C2H2-TYPE 5.
FT ZN_FING 492 514 C2H2-TYPE 6.
FT VARSP:IC 10 53 Missing (in isoform Ik4).
FT VARSP:IC 54 140 Missing (in isoform Ik2).
FT VARSP:IC 54 283 Missing (in isoform Ik6).
FT VARSP:IC 197 283 Missing (in isoform Ik3 and isoform Ik4).
FT VARSP:IC 197 238 Missing (in isoform Ik7).
FT VARSP:IC 141 283 Missing (in isoform Ik5).
FT CONFLICT 11 12 QV -> PS (IN REF. 2).
FT CONFLICT 214 214 S -> T (IN REF. 2).
FT CONFLICT 245 245 N -> X (IN REF. 2).
FT CONFLICT 296 296 MISSING (IN REF. 2).
FT CONFLICT 298 298 S -> T (IN REF. 2).
FT CONFLICT 352 355 KPLA -> RRS (IN REF. 2).
FT CONFLICT 372 372 N -> Y (IN REF. 2).
FT CONFLICT 420 426 PHARNGCL -> RRAQRY (IN REF. 2).
SQ SEQUENCE 519 AA; 57528 MW; 7B0129C4E3F41A8 CRC64;
Query Match 3.9%; Score 269; DB 1; Length 519;
Best Local Similarity 25.9%; Pred. No. 4; e-08;
Matches 139; Conservative 67; Mismatches 192; Indels 138; Gaps 29;
QY 614 LTPVSGTQGTQTLRESNPVAAHARIPIYGSVATKTFLOGAPA---GGEKSGALPQGY 87C
DB 37 LSTTSGQGSSK-----SDRVASN-----VKVETQDEENGCRACENGBEACEDJRL 95
QY 671 PASGENKS---KDESCLLRRRGSGV-----FCA-NCLTTKTLNKKANG 913
DB 96 DASGEKNGSHRQGGSSAL-----SGVGIR;PNQKAKCDICGICIGPNVVMVHRSH 140
QY 914 G---YYCNACGL-----YQKLHSTPRPLNIKQNGEQIIRRRTRKLNPEALQA 960
DB 141 GERPFQCNCGASFTCKGNLLRHILHSGEKP---FKCHLCNYACRRR-----DALTG 190
QY 961 EQLNKQQRSGNEEQNGSPLRRS---EDHLTESHO--REIPLPS-----LSKYEAQGS--- 1009
DB 192 HLRTHSVGRKPHKCGYGRSYKQRSLEBHKRCHRYLESMLGPTLYPVYKKEETHSEMA 250
QY 1010 --LTKSHSACQPVLSQTLDIHKRMQPLHIQKSPQESTGDPGNSSSVSEKSGSSESGSP 1067
DB 251 EDLCKTGSERSILVRLASNVAKRYSK-----PQKELGKGLSDTPYDSASVEKENE 304
QY 1068 IEK-----YVRPAKHPNVSPG-----SPIEKYQYPLF-GLPFFV 1100
DB 305 MKKSHVMDCAINNAINYLGAESLRPLVQ---TPPGGSEVWPVISPMYQLHKPLAECTFRS 361
QY 1101 HKDFQSEA-DKLRFWSKYKLSVFGNPHYLSHVFGPL-NPCQNYVPYPTFNLPPHPSAVGS 1158
DB 362 NHAQDSAVENLLJLSKAKL-VF-----SEREASPSNSCQDSTTESNKEEQRSGLIYL 414
QY 1159 DNDIPLDLAIKHSRPGPTANGAS-KETKA-----PNNVKNEGPLNVVKTKEVDSTODE 1212
DB 415 TNHI-----APEAR-----NGLSLKEHRAYDLLRAASENSODALRVVST-----SGSQM 459
QY 1213 LSTKVKHCGIVFJDEVYVALHMSCHDGGPFOCSICQHLCTDKYDFTTHIQRGLHR 1268
DB 460 KVKKCEHCRVLF-DHWYV-TIEMGCHGRPPFECNVCYGVHSQRYEFSHH-IRGEHR 515

```

RESULT 7
HUNB DROR STANDARD; PRT; 767 AA.
AC O62537;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hunchback protein.
GN HB.
OS Drosophila orena (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7233;
RN [1]
RP SEQUENCE FROM N.A.
RA Tautz D., Nigro L.;
RT "Microevolutionary divergence pattern of the segmentation gene
RT hunchback in Drosophila.";
RL Mol. Biol. Evol. 15:1403-1411(1998).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ005375; CAAG6505.1; -
DR FlyBase; FBgn0024421; Dore\bb.
DR InterPro; IPR007087; Znf.C2H2.
DR Pfam; PF00096; zf-C2H2; 6.
DR SMART; SM00355; Znf.C2H2; 6.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW Developmental protein; Gap protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 61 71 POLY-GLN.
FT DOMAIN 184 210 POLY-GLN.
FT DOMAIN 119 224 POLY-HIS.
FT ZN_FING 242 264 C2H2-TYPE.
FT ZN_FING 271 293 C2H2-TYPE.
FT ZN_FING 299 321 C2H2-TYPE.
FT ZN_FING 327 351 C2H2-TYPE.
FT ZN_FING 410 415 POLY-GLN.
FT DOMAIN 530 533 POLY-GLU.
FT DOMAIN 686 693 POLY-SER.
FT DOMAIN 698 703 POLY-ALA.
FT ZN_FING 714 736 C2H2-TYPE.
FT ZN_FING 742 766 C2H2-TYPE.
SQ SEQUENCE 767 AA; 84207 MW; 9C7710B7AED78914 CRC64;

Query Match 3.9%; Score 266.5; DB 1; Length 767;
Best Local Similarity 19.0%; Pred. No. 9.7e-08;
Matches 170; Conservative 116; Mismatches 354; Indels 257; Gaps 30;

QY 472 NDKLSRGVINDLAKSSEGETMTYTKSSGAKKDFSSKGAEDNMVT-----SY 523
DB 26 NIKOEPGHLGDSNVASSPQSPSTNHLEQFLKQQQQHQHQQPMDTLCAMTSPFSQ 85
QY 524 NCQCFDFRYKSGHPDIVVUGPLLRHY----QQLNHIKCTIKGPFPCPRGLCSPEKH-- 577
DB 86 NDQNSLQHYDANLQQGLLQQQQQQOQHQFQAQQOHHHHHLMGCFNPLTPPG.PNPMCHFY 145
QY 578 ---LGEITVPFACRKNCSHCALLLLHLSPGAGSGSRVKEQCHQCSFTTPDQVLLFHYE 634
DB 146 GGNVLRSPQTPISASTVASVAV-----ATGSS-----EKQALTPPKDVT--PPK 189

```

```

QY 635 SVHESQASVYKQFAHRLQSGDGGQSVK---ESKEH----- 666
DB 190 SPAKSSQSNIEPEKEHEDQMSNSEDKMYMVESEDDTINIMP.YNSHGKMKVYKCKTCGV 249
QY 667 -----SCTKCDFTQVEEISISHYR--RAHSCYKCCQCSFTAA 702
DB 250 VAITKVDFAHTRHMKDPDKILOCPKCPVFTEFKHLEHYH-RKHKKQKPFQCDKCSYTCV 309
QY 703 DTCSLLHFHNT-----VHCQEDDIT-----ANGEEDGHAISTIKER-----PKIDF 744
DB 310 NKSMKSHRSKSHSVYQYRCADYATKYCHSEKLNKRYCHKPGMWLDEGTNPSLVI 369
QY 745 RYVNLCT-PCKYGEVPVSESVKLEEKGLKEXVWTESSDDLRNVWARGADLRGS 833
DB 370 DVTGTRRGPKSKAGGPIASG-----GS-----GS 393
QY 804 PSYTOASGLLTPTVSGTQECTKT.RDSPNVAAHARPIYGLAVEYKGF--LOGAPAGGE 961
DB 394 GSGSGSRKSNVAAPQCCQCTQPTOP-----P-SQLSAAJGGFPLVQSNAPPA 442
QY 862 KSGALPCQYPASGENSKDESQSLRRRRSGVFCANCLTTKTSJMRKNANGGVVNCACG 921
DB 443 ASPLLPLFVSPAKSVASVEQTPLSP-----ANKLPPLASLLQCNHMAFF----- 489
QY 922 LYOKLHSTPRLNIKKNGEQI--RRTRKRLNPEALQAEQNKQORGSEEQVNGSPLE 981
DB 490 PYWNLN-----LQMLAAQQQAAVL-----AQLSPRYRECLQQNQCCSDNEEEQDEYE 539
QY 982 RRSDDHUTESHQREIP-PSLSKYAEAGSLTKSHSAQPVLSQTLCTLCK-HKE--MQPLHIQI 1039
DB 540 RKSVDSSAMCL-----SQGTPVKEEQQ-----LHQCPQQLVMNL 576
QY 1040 KSPQESTGDFGNSSSVSEKSGSSRSGSP-IEKYNRPKAPKPNYSPGSPIEKYQVFLGLPF 1099
DB 577 KVEEAT-PLVSSSNARRKG---RVLKJDLTLQLRSGVWTSPE----- 616
QY 1100 VHNDFOSEADLFWSKYKLSVFGNPHYLSHVPLGNLPCNQYVYPYPTFNLPFHSAVGSD 1159
DB 617 -----QLKVFSTP-----MPTASSPIAGRKEMP-----EDHCSGTSSA 649
QY 1160 NDIPDLAIKH-----SRPGPTANGASKETKAPPNVKNEGPLNVYKTEKVDNSTQ 1210
DB 650 DESNETAHRQCANTSSASTASSGNSGNSNNGSSSSSSSGTNSAAAAA--PSGTAA 709
QY 1211 DELSTKCHVCGIVFJBEWYALHMSCHGSGPFCSCICQHLCTCKYDFTHIQRLH 1267
DB 710 AGAIVECKYCTIFFKDAVLVTHMGVHSCDVFKNCMGCKGCDGPGVGLFVHVARNAF 766

RESULT 8
Y296_HUMAN STANDARD; PRT; 1829 AA.
AC O15015;
DT 16-OCT-2001 (Rel. 42, Created)
DT 16-OCT-2001 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical zinc finger protein KIAA0296.
GN KIAA0296.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
EX XEDLINE=9734998; PubMed=9205941;
RA Nagase T., Ishikawa K.-I., Nakajima D., Chira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 103 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL CNA Res. 4:141-150(1997).

```

CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB002294; BAA20756.1; -;
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; Zf-C2H2; 28.
DR SMART; SMO0355; Znf C2H2; 31.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE; PS00557; ZINC_FINGER_C2H2_2; 29.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT ZN FING 8 31
FT C2H2-TYPE.
FT ZN FING 48 70
FT C2H2-TYPE.
FT ZN FING 75 97
FT C2H2-TYPE.
FT ZN FING 239 261
FT C2H2-TYPE.
FT ZN FING 266 288
FT C2H2-TYPE.
FT ZN FING 294 316
FT C2H2-TYPE.
FT ZN FING 374 396
FT C2H2-TYPE.
FT ZN FING 401 424
FT C2H2-TYPE.
FT ZN FING 465 487
FT C2H2-TYPE.
FT ZN FING 492 514
FT C2H2-TYPE.
FT ZN FING 575 597
FT C2H2-TYPE.
FT ZN FING 617 639
FT C2H2-TYPE.
FT ZN FING 644 666
FT C2H2-TYPE.
FT ZN FING 821 843
FT C2H2-TYPE.
FT ZN FING 848 870
FT C2H2-TYPE.
FT ZN FING 881 904
FT C2H2-TYPE.
FT ZN FING 958 980
FT C2H2-TYPE (DEGENERATE).
FT ZN FING 1052 1074
FT C2H2-TYPE.
FT ZN FING 1079 1101
FT C2H2-TYPE.
FT ZN FING 1203 1225
FT C2H2-TYPE.
FT ZN FING 1230 1252
FT C2H2-TYPE.
FT ZN FING 1258 1280
FT C2H2-TYPE.
FT ZN FING 1299 1321
FT C2H2-TYPE.
FT ZN FING 1326 1348
FT C2H2-TYPE.
FT ZN FING 1364 1386
FT C2H2-TYPE.
FT ZN FING 1557 1579
FT C2H2-TYPE.
FT ZN FING 1585 1607
FT C2H2-TYPE.
FT ZN FING 1677 1699
FT C2H2-TYPE.
FT ZN FING 1704 1726
FT C2H2-TYPE.
FT ZN FING 1732 1754
FT C2H2-TYPE.
FT ZN FING 1761 1783
FT C2H2-TYPE.
SQ SEQUENCE 1829 AA; 200760 MW; B20C482B38684895 CRC64;

Query Match 3.7%; Score 251.5; DB 1; Length 1829;
Best Local Similarity 18.4%; Pred. No. 2.3e-06;
Matches 279; Conservative 159; Mismatches 485; Indels 593; Gaps 77;

QY 56 KHEKSLHVQDPSSSSKKD-LKSAVLSEKAGFNYESPSKGGNFFSPHDEVTDRNMLA-- 112
DB 26 HRE--LLHPSNQDSEADSIIPRYCQCQCGYRHPGSLN-----HRRTHEGLFPCT 78
QY 113 -----FSPPAGVCVEPLKSPORABD-----PQDMACTPS-----GSGSJE 149
DB 79 TCCKDFSNPMA-----LKSHMRTHAPEGRRRRHRPRPKEATPHLGQETVSTQMGQRLG 132
QY 150 TKEDQMSKATEETQQAQS-----GCAN-----CQGLSPVSV 182
DB 133 SSEGMENQTKTEETPDCESPDPPRAAGTWEDJPTROREGSLASHPGPDGADGMPSTN 192
QY 183 ASKNQVPSDGGVRLN-----KSKTLLVNDNDPAPLSPQLQDFKNCICG--YGVYG 233
DB 193 SARAPLPPIPASSILLSNLEQYLAEVSNVFTGQEPQSPFAEERRYKSCGCKTYRHAG 252

QY 234 NDPTOLIKFRKRYHLGLHNRTQDAELDSKILALHNMVQFQSHKDFOKVNRSVFSGVLQD 293
DB 253 S-----LUNHRQSHTLGIYPCAICFKPEF-SNLMALKN-----HSR----- 286
QY 294 INSSRPVLLNGTYQVQVSGGTFIGIGRKTDCQGTQNTKYFRCKFCNFTYMGNSSTELQCH 353
DB 287 -----LHAGY-----RPHCHPCPRVF--RLPDELLEH 312
QY 354 FLOTH-----ENKIKASLPSEVAKPKSEKSNKSI-PALQS-----SDSGELGKWQCK 400
DB 313 -CQSHGEGEQERWEKEMPTTN--GHIDESQQLPSAQMNLNGSAELSTGEL----- 363
QY 401 ITVXAGDQTPVGVYVPIKPLDSSR-----QNGTEATSYVMCKFCSCFSSSSSJK 450
DB 364 -----EDSGLEBYRPFPGDQGRYRHAGSLNHRKSHQGVYPCSLCSKQLFNAALK 417
QY 451 LLEHYGKHQGVAVCGSLNPELNCKLSRGSVINQNDLAKSSEGETNKTCKSSSGAKKQDF 510
DB 419 --NVRARHRRPPQGVGENGQ--PSYPPAPLLAEATHKEEEDPTTLCHR----- 463
QY 511 SSKGAEDNMVTSYNGQCFDFRYSKSHGPDVIVUGPLLRHYQCLNHRKCTIKKCFPCERG 570
DB 464 -----PYKCEGCRAYRHR-----SSLVNH-----RSHRTGEVCCSCJCR- 499
QY 571 LSSPEKHLGETYP--FACRKSNCGHC-----ALLLHLSPOGAAGSSRYK 613
DB 500 -----KYPNLVALRNHVRYHCKAARRSACI-GAEGAFSHLKVLEFPDPVEAERP 549
QY 614 H--QHCCSFTTPOVTVLLPHYESVHESQASVYKAEHJQSSDQGOOSVKESKEHSCYCK 671
DB 549 HTDQCHVCK-----HEEATDITPAAD-----KTAAHICISIC 580
QY 672 DFTOVEEESIRH-----YBRAHS-----CYKROCSTAAADTQSLLEFNTVH 715
DB 581 GULLFEDASSLERHGLTHGAGEKENSTETMPPAPACRDCGKSYRHSGLNHRCT-H 539
QY 716 CQEQITTTANGEEGHAISTIKKEPKIDFRVYNLITPQSKMGEVSVSVKREKJEEKDG 775
DB 640 -QTGDFSCGACAKRPHITMAAK-----NHLRRHSR-----RRSRHRKRA 678
QY 776 -----LKERVWTSSESSDDLRNTWRGADILRQSPSYTQASLGILLTPVSGTQFCT 824
DB 679 GGASGGREAKLLAESWTRELD-----NEGLES----- 738
QY 825 KTLRDSPNVEAAHLARPIYGLAVETKGLQAGAGKSGALPQYVPSGKNSKDESOS 884
DB 709 ----QDPSSGESPHGEG-----NLESDGCLQAESEGCCKGLERDETHFGDKESGCTGEG 760
QY 885 LLAR-----RRSGV-FC-----ANCLTTKTSLWR 908
DB 761 LERKDAASLLDNLIDFQEEGGTHFCDSLTGVDGQKATGPQNPSSSHSANAVTG---MQ 816
QY 939 KWANGOVVNCAGLYOKLHSTPRP-LNII-----KQNGEQIIRRRTRK 951
DB 817 --AGAAHTCSDCG-----HSFPFATGLLSHRCHPDPGIVCCSLCPKFPDLPALASHFON 869
QY 952 RLNPALQAE-----QLNKQ-----RGSNEE-CVNG----- 977
DB 870 HRPGEATSACPF-CCLCGMIFPGRAGYLRHRQAHSSSGMTGSEEGEEGVAEAAAP 929
QY 978 SPLERRSEDLTHESOREI-PLPSL-----SKYEAQSLTKSHSAQCPVLVSQTL 1026
DB 930 SPFLQ-SEAEJLNQLOREVEA-CSAGYGHICCCQCTYDGLGSLRRHHQSQ-----SSGT 984
QY 1027 D-HKRNQPLHI-----CIKSPCESTGDP-GNSSSVSESGK 1060
DB 985 TACKAPSLGVAGAGANEXVVDVULEDI-VNSVSEGGDAKS-QEGAGTPUGDSJC-CQGG- 1042
QY 1061 SBERGSFIKKYVRPAK-----HFNYSPPGSPIEKYCPJFGLPFVHNDQFQSEACWL----- 1111
DB 1043 ----SLLEAQPRPCNCGKTYRHGSLVN-----HRKIHTQTCFLCPVCS 1085

```

QY 1112 RFWSYKLSVPGNPHYLHVPLPQPCQVYPTFNLPHPFSAVGSNDIPLDJAIAKHS 1171
DQ 1086 RCY-----PFLAAYRNHLRHPR-CRGSPE-----QVCPPEAAGSS-----ELCV--- 1125
QY 1172 RPTGANGAKETKAPPNVKNKP-----LAV-----VET 1202
DE 1126 --GPIPEGSNK-----PQMAEEGQQAQVEKQLQELKVEPLEEVARVKEEWEETTVMG 1179
QY 1203 EKV-----DRSTDELSTK-----CVHCGIVFLDEVMYALHMSCHGSGPFGQCSICQHL 1251
DQ 1180 EIEPRLEJAEKGCQTEASSRPFCEVCGRSYKHAAGSLINRQSH-CTGHEGCAACSKG 1236
QY 1252 CTDKYDFTHIQGLH 1267
DQ 1239 FSNLSMLKNH--RRTH 1252

RESULT 9
HUNB_DROYA
ID HUNB_DROYA STANDARD; PRT; 759 AA.
AC 062541;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-SEP-2003 (Rel. 42, Last annotation update);
DE Hunchback protein.
GN HB.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
SEQUENCE FROM N.A.
RA Tautz D., Nigro L.;
RT "Microevolutionary divergence pattern of the segmentation gene
hunchback in Drosophila.";
RL Mol. Biol. Evol. 15:1403-1411(1998).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
OF HEAD STRUCTURES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ005376; CAA06506.1; -
DR FlyBase; FBgn0022824; Dyak\hb.
DR InterPro; IPR037087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 6.
DR SMART; SM00355; ZNF_C2H2; 6.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW Developmental protein; Gap protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 61 70 POLY-GLN.
FT DOMAIN 103 109 POLY-GLN.
FT DOMAIN 118 123 POLY-HIS.
FT ZN_FING 241 263 C2H2-TYPE.
FT ZN_FING 270 292 C2H2-TYPE.
FT ZN_FING 298 320 C2H2-TYPE.
FT ZN_FING 326 350 C2H2-TYPE.
FT DOMAIN 405 410 POLY-GLN.
FT DOMAIN 525 528 POLY-GLU.
FT DOMAIN 555 564 POLY-GLN.
FT DOMAIN 690 695 POLY-ALA.
FT ZN_FING 706 728 C2H2-TYPE.
FT ZN_FING 734 758 C2H2-TYPE.

```

```

SQ SEQUENCE 759 AA; 93287 MW; 67D37A252978D365 CRC64;
Query Match
Best Local Similarity 19.3%; Score 248.5; DB 1; Length 759;
Matches 174; Conservative 11; Mismatches 344; Indels 271; Gaps 36;
QY 472 NDKLSRGSVINQNDLAKSSEGETMTYTDKSSGAKKCFSSKGAENMVT-----SY 523
DQ 26 NIKCEPCHHEDGNSVASSERQSPISPTNHLCEPLKQQQ-QQQHQQQPMQDTCAYTRSPSQ 84
QY 524 NCQCFDFRYSKSGPDIIVVGPPLRLHY-----CQJHNIHKCTIKHCPCPCRGCLSPKHK-- 577
DQ 55 NQONSICHYDASIQQLQQLQQQYQCHQFQAQQQHHLHLMGCFNPLTPGLPNMWHFY 144
QY 578 ---LGEITVFACRKGNCSCHCALLLHLSPGAAGSSRVZKHQCQCSFTTDPDVLFLHYE 634
DQ 145 GGNLRPSQPTPTASSTVAPVAV-----ATGSS-----EKLCAITFPMDVT--PPK 188
QY 635 SVHESQASDVQKQANHLQSGDQGSVK-----ESKEH----- 666
DQ 189 SPAKSSQSNTEPEKEHCHQNSNSEDMMKYMAESEDDETNIKYP-YNSHGKVKYKCKTCGV 248
QY 667 -----SCTKCDPITOVEEELSRHYR--RAHSCYKCYCQCSFTAA 752
DQ 249 VAKYDFWATHTHMKPKILQCPKCPVTEFKHLSYHIRKHKQKQPCQCDKCSFTCV 328
QY 703 DTOSLLEHFN-----VHCQEQQIT-----ANGEEDGHAISTIKEE-----BKIDF 744
DQ 309 NKSMVSLNHRKSHSVYCYRCACQDQATKYCHSPKHLRKYGHKPGMYLDEGTPNPSLVI 368
QY 745 RVYNLLT-PDSXGEPVSESVVREKLEEKDGKLEKWTESDDJLRNVTRQADILRGS 803
DQ 369 CVYGTTRGPKSKNGGPIASG-----GSGSGS-----RKPNAAVA 423
QY 804 PSYTOASLGLLTPVS-GTCECTKTLRDSNVE---AAHLARPIYGLAVETKGLQAGAPAG 859
DQ 404 PQCCQSQ-----PAQPATLSLSALQGPFLVQNSAPPASPVLP-----PAS 447
QY 860 GEKSGALPQQVPASGENKSKDESQSLRRRGRSGVFCANCLTTKTSLRKNANGGVYCN 919
DQ 448 PAKSVASVEQTPLSPF-----ANLLPPLASLQQNRKYAFF--- 484
QY 920 CGLYQKLHSTFRPNJ::KQNGEQIIRRTKRLNFEALQAEOLNKCQSGSNEEQVNGSP 979
DQ 485 --PYWNLN-----QMLLAQQQAAVJ-----AQLSPMRSEQLQCCNCHSDNEEBEQDDE 532
QY 980 LERPSDELHTEHQREIPLPFLSKYCAQGLTKSHSAQOPVLVSQTLD-HKRMCPHLHIQI 1039
DQ 533 YERKSVDSAVQLSQ-GTPVKEDDQHQQQ-----QQP-----CQPLAKNL 570
QY 1040 KSPQESTGDFGNSSVSESGKSSERSGPIEKYVRPAKHPNYPSPGPIEKYQYPLGLPF 1099
DQ 571 KVEEAT-PLMSSSNASRRKG---RVKLQTLQ----- 600
QY 1100 VHNDFQSEADWLRFWSKYKLSVPCNHYLSHVPLGJNFCQNVYVPTFNLPHFSAVGS 1159
DQ 601 -----LRSEA-----MTSPCLKVPSTP-----MPTASSPIAGRKPMF-----EDHCSQTSSA 643
QY 1160 ND-PLDLAIAKH-----SRPGPTANGASKKEKTKAPPNVKNYEGP-VVVKTEKVD 1207
DQ 644 DE--SMETHAVACANT-SASSTASSGNSGNASSNGNSS--NSSNGTSSAAAAPASGT 699
QY 1208 STODELSTKCVHCGVFLDEVMYALHMSCHGSGPFGQCSICQHLCTDKYDFTTHIQGLH 1267
DQ 699 PAAAGAIYECKYCDIFFKQAVLYTHVGYHSCDEVPKCYCGEKCDGPGVGLFVHYNAAH 758

RESULT 10
TCF8-CHICK
ID TCF8-CHICK STANDARD; PRT; 1114 AA.
AC P36137; C42408;
DT CL-JUN-1994 (Rel. 29, Created)
DT 31-JUN-1994 (Rel. 29, Last sequence update)

```



```

QY 1185 --TKAPPNVKNEGLNVVYKTEKVDSTODELSTKCVHGVFLDEVVYALHMSCHGDSGP 1242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 873 TSSEGVSNVEDONSDSDTPPKKQKRTKNGMYA-CDLCKIFQKSSLLRHKVHTKRP 93:
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1243 FQCSICQHLCTDKYDFTTHIQGLRNNAQVEKNGK 1278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 932 HECGICKAFKHKHHLIEHMLHSEKPKYQCDKCGK 967
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
ABFL1_HUMAN
ID ABFL1_HUMAN STANDARD; PRT: 3703 AA.
AC Q15911; O15101; Q13719;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE (AT-binding transcription factor 1).
GN ATBF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM A).
RP TISSUE=Lung;
RX MEDLINE=96070776; PubMed=7592926;
RA Miura Y., Tam T., Ido A., Morinaga T., Miki T., Hashimoto T.,
RA Tamaoki T.;
RT "Cloning and characterization of an ATBF1 isoform that expresses in a
RT neuronal differentiation-dependent manner.";
RL J. Biol. Chem. 270:26840-26848(1995).
RN [2]
SEQUENCE FROM N.A. (ISOFORM B).
RX TISSUE=Hepatoma;
RC MEDLINE=92049333; PubMed=1719379;
RA Morinaga T., Yasuda H., Higashio K., Tamaoki T.;
RT "A human alpha-fetoprotein enhancer-binding protein, ATBF1, contains
RT four homeodomains and seventeen zinc fingers.";
RL Mol. Cell. Biol. 11:6041-6049(1991).
RN [3]
SEQUENCE OF 1-1190 FROM N.A. (ISOFORM A).
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Shedd V.P., Kalush P., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barstead M., Cronin J.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN [4]
SEQUENCE OF 1151-3703 FROM N.A.
RA Kozlowski A., McQuerry Y., Hotic M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC sequence of the enhancer element of the AFP gene.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q15911-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q15911-2; Sequence=VSP_006825;
CC -!- SIMILARITY: Contains 4 homeobox domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L32832; AAC14462.1; -
DR EMBL: D10250; BAAC1095.1; -
DR EMBL: AC002044; AAC31674.1; -
DR EMBL: AC004943; AAC79153.1; -
DR HSSP: P20263; 10CP; -
DR TRANSPAC; T00048; -
DR TRANSPAC; T01665; -
DR Genes; HGNC:777; ATBF1.
DR MIM; 104155; -
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:003705; F:RNA polymerase II transcription factor acti...; TAS.
DR GO; GO:006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf-C2H2.
DR Pfam; PF00046; homeobox_4.
DR Pfam; PF00096; zf-C2H2_18.
DR PRODOM; PDC0010; Homeobox_4.
DR PROSITE; PSC0027; HOMEBOX 1; 2.
DR PROSITE; PSC0071; HOMEBOX 2; 4.
DR PROSITE; PSC0028; ZINC FINGER C2H2 1; 14.
DR PROSITE; PSC0157; ZINC FINGER C2H2 2; 9.
DR Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Homeobox; Nuclear protein; Repeat; Polymorphism;
KW Alternative splicing.
FT ZN_FING 282 305 C2H2-TYPE.
FT ZN_FING 640 663 C2H2-TYPE.
FT ZN_FING 671 694 C2H2-TYPE.
FT ZN_FING 726 750 C2H2-TYPE.
FT ZN_FING 804 828 C2H2-TYPE (ATYPICAL).
FT ZN_FING 945 968 C2H2-TYPE (DEGENERATE).
FT ZN_FING 984 1009 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1040 1064 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1223 1246 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1252 1275 C2H2-TYPE.
FT ZN_FING 1360 1385 C2H2-TYPE.
FT ZN_FING 1401 1423 C2H2-TYPE.
FT ZN_FING 1429 1452 C2H2-TYPE.
FT ZN_FING 1545 1569 C2H2-TYPE.
FT ZN_FING 1596 1620 C2H2-TYPE.
FT ZN_FING 1993 2026 C2H2-TYPE.
FT DNA_BIND 2165 2204 HOMEBOX 1.
FT DNA_BIND 2442 2501 HOMEBOX 2.
FT ZN_FING 2328 2351 C2H2-TYPE (ATYP-CAL).
FT ZN_FING 2530 2552 C2H2-TYPE.
FT DNA_BIND 2641 2700 HOMEBOX 3.
FT ZN_FING 2711 2734 C2H2-TYPE.
FT DNA_BIND 2944 3003 HOMEBOX 4.
FT ZN_FING 3024 3048 C2H2-TYPE.
FT ZN_FING 3529 3553 C2H2-TYPE.
FT DOMAIN 104 107 POLY-PRO.
FT DOMAIN 460 489 POLY-GLU.
FT DOMAIN 770 784 POLY-ALA.
FT DOMAIN 1723 1743 POLY-GLN.
FT DOMAIN 1789 1794 POLY-GLN.
FT DOMAIN 1952 1957 POLY-GLN.
FT DOMAIN 2337 2352 POLY-PRO.
FT DOMAIN 3197 3209 POLY-GLN.
FT DOMAIN 3210 3214 POLY-PRO.
FT DOMAIN 3231 3231 POLY-GLN.
FT DOMAIN 3376 3389 POLY-GLN.
FT DOMAIN 3392 3395 POLY-GLY.
FT DOMAIN 3507 3527 POLY-PRO.
FT DOMAIN 3597 3600 POLY-SER.
FT VARSPLC 1 914 Missing (in isoform B).
FT VARSPLC 3374 3374 A -> V.
FT VARSPLC 3377 3384 MISSING.
FT VARSPLC 3527 3527 G -> GGG.

```



```

FT CONFLICT 72 72 /FTID=VAR_011696
FT CONFLICT 422 422 A -> P (IN REF. 3).
FT CONFLICT 579 579 A -> A (IN REF. 3).
FT CONFLICT 767 767 I -> S (IN REF. 3).
FT CONFLICT 777 777 A -> V (IN REF. 3).
FT CONFLICT 846 849 HHRV -> RHLG (IN REF. 3).
FT CONFLICT 997 997 A -> S (IN REF. 3).
FT CONFLICT 1150 1190 EEAIEVDEGSEPTADPESLAKDQGGGASSOAEKELTSP
FT -> GEWSHRGRPRGLGVHLLSTSRGLLFEQVYCPAGPH
FT VPY (IN REF. 3).
SQ SEQUENCE 3703 AA; 404468 MW; 0562AF37D4DCF856 CRC64;

Query Match 3.6%; Score 245; DB 1; Length 3703;
Best Local Similarity 17.8%; Pred. No. 1.4e-05;
Matches 292; Conservative 214; Mismatches 515; Indels 62; Gaps 80;

QY 22 EPIGTESKVGKKEFSADQMSNTQSDAAELNHKEE-----HSLHV----- 64
DB 575 EGVRTNVAEGRRLLDF-ADP-SANKONATAPERNSTEGDGGFVPHHGHAGSLCELGVG 632
QY 65 QDPSSSKDKL-K-SAVISEKAGNYEPPSKGNFP-----SFPHEV 105
DB 633 ECPSSGVECPKCDTVLG-----SSRLGGHMTVGHSRNSCKTLKCPKCMHXYKQOT 685
QY 106 TDRNMLAFSPAAAGVCEPLKS-----PQRAEAD-----DPQDMACTPSCG-SLE 145
DB 686 LEAHMKE-KHPEPGGVYCKSQOPHPR-LARGSYTCGVKPFRCVNCVYTTTKGNLSIH 744
QY 150 TKEDQMSKATETQAGSQANCOG--LSPVSVASKNPQVPDSQGVRLNKSCTELLVN 237
DB 745 MQSDKLNHYNQNLQGGGQVFIHTAGAAAAAANAANSSSCGA----- 792
QY 208 DNEPAPLSPLO-DPKNICIGVYGYNDPTDLKHFR-----KVHLG----- 249
DB 793 ----PSPTKTKPTWRCEVDY-----ETVARNLRIMHSEKHMNMMLQGNMTQI 842
QY 250 LHNRT-----QDAELDSKILA-----LHNMVQFHSKDPQVNRSVFSGVLCD----- 293
DB 843 QNHHRVLGSLPAPAEALYYVLAQNMNLPNLKMDSAASDAQ-----FMWSGFQDPA 898
QY 294 INSSRPVLNGTYDVQVTSG-----TFIGIGR-----KTPDCQGNTKYFRCKFN-FT--- 341
DB 899 MAATPALVGEIPLDMRLGGGLVSEELMNLGSEFIQND--PSLXFFQCAVCNKFITD 956
QY 342 ----YNGNS-----STELQHFLOTHPNKIKASLPSEV 371
DB 957 NLDMLGLHMVRSLSDEWKAVMGDSYQCKLCRYNTQLKANP-QLHCK----- 1004
QY 372 AKPSEKNSKSI-PALOSSDSGLKQWQKITVAGDDTFVGYSVPIK-----PLDS 422
DB 1005 ---TDRHVQYQLVAHIKEGKANERLKCVA-----IGNPVHLKNCADYITNSLEK 1054
QY 423 SR-----QNGTEATS-YVMCKFCFSCESSSLKLEHY-GKH 459
DB 1055 LRJHTVNSRHEASLKYHLQCHESGVGESCYHCVLCNYS---TKAKLNL-CHVRSMKH 1112
QY 460 GAVQS-----GGLNPELNDKLSRGSVINQ-----NDLAK 488
DB 1113 QRSESLKRLQRLQGL-PEDEDLQGITFIRRCPTSTCPBEA-EDVEGPSETAADPEELAK 1171
QY 489 SSEG---ETWTKTDKSSSGAKKDFSSKGAEDNMVTS-----YKCOFCQFR 531
DB 1172 QDEGASSQAELTDSPTSXRI-SPPGSSSEPLSKAPKTAEE-KPQMYQCTPYC--K 1229
QY 532 YKSHGPDVIVWGLLRHYQQLNHNKKT-IKHGPFPCPRGLCSPKXHLGSEITYFACRKSN 591
DB 1230 YSNAD-----VNLRLVHANTQHSVQ---PMLRCP-LCCMX-----NW 1263
QY 592 CSHCALLLHLSPGAGSSRVKHQCHQ---CSTTTPDVV---LJPHVESVHESQASDVQK 646
DB 1264 KIHQLHQLTHL-----HSVAPDCVEKLIMTIVTTPMVVPSMFLPAAVPPDRGKSNLE 1316

```

```

QY 647 EA-----NHLCQSCGQ-----SVKSEKHSCTK--CDFITQVEEE 680
DB 1317 EAGQKQETSEDLGKN:LPSASTEQSGLKPSPADPGSVREDSGFIWKRKGNQVFTSAA 1376
QY 681 ISRHVRAHS-----CYKRCQCSPT-----AACT----- 704
DB 1377 LQTHFNEVAKRQPLQFVSRHVKVYRCNCSLAFKTIKJLQHSQYHVRANTMCLQCR 1436
QY 705 -----GSLLEHNTVHCQ--EODI-----TTANBEEQHA:STIKSEPKIDFRVYNLLT 751
DB 1437 SFTTFQALAKHLETSHLE:SEADIQOLYGGLLANGD-----LLA 1475
QY 752 PDSKMGEPV---SESVVXREKLEEKDGKLEK-----VWTESSSDDLNENWRCAG 798
DB 1476 ----MGDFLAEDHIIIEEDKEESDLEDKSPGSSGSGVQEDSGSEPRALPFR--- 1529
QY 799 LDRG-----SPSYTOASLGLL--TPVSGTQEQTKTJLDRSFNVFAA 836
DB 1529 --KGNFTYKXP-LPSEPVKCTVCKSESFTQKNILLVHYNSVSHJHLKRALQESATGOPE 1586
QY 837 HLARP-----YQ-----LAVETKQLOQAPAGGKSGALPQOYPASGENKSKD 880
DB 1587 PTPSPONKPFKNTCKVAYSQSSSTLEIHRPSVLDHQTAKAAKLEA-----ASGSSNQTG 1640
QY 861 ESQSL-LRRRRSGGVPCANCLTKTSLMRKNANGGVVNCACGLYQKLHS-----TP 930
DB 1641 NSSSISLSSTSPVSTSGSNTFTTS---NPSSAGIAPESNLLSOVPTEVGMPLGNP 1696
QY 931 KPLNLT-KQNGNFCITRRTRKRLNFEALCABQJMKQCGSNEECVNGSPLEERS--EDHL 988
DB 1697 IGAKEASPSEPREANRKLADVIASRQCCQCCQCCQCCQCCQCCQCCQCCQCCQCCQ 1756
QY 989 CESHORELP-----PSLSKYEAQGLTKSHSACQPVV-----SOTLDIHKR--- 1031
DB 1757 QCELOQCALIQSOLENPTLLPHFPXTETJLQLOQOQHLFFPYPGAEFCNLNEVSLP 1816
QY 1032 -----YQPLH:QIKSPQES-----TQCPNCSVSVSEKSGSSRSPLEKY 1071
DB 1817 VTSGALTLTGTPGLLEDLKAQVQVPOQSHQQLPQOQCNLSIAQSHSA-----L 1867
QY 1072 MRAPKPNYSPGSPLEKYQYPLGLPVHDFQSEADLRFWSKYKLSVPCNPHYLSHV 1131
DB 1868 LQSPQKPKKKNKLVYKKE-----EKESQERD----- 1894
QY 1132 PGLNPNQNVYPTFNLCPHPSAVGSDNDI-----PLDLAIKHSPGPTANGASKETKA 1187
DB 1895 -----SAGEGEGTGPKEITLPDALKAKEKKE:APGGSGSPSVL 1932
QY 1188 PNVKNEGPJNVVKT-----EKVDRSTQDEJSTK--CVHCGVVF 1224
DB 1933 PPIASDARGNATKALLENFGFELVIQVNNENKQVCKXNGKTQDGENLEKUECDSCGKLF 1992
QY 1225 LDEWYVALHYS-CHSDSGPFO 1244
DB 1993 SNILLKLSQEHVHQNYFPFK 2013

RESULT 12
GAT2 XENLA
ID -GAT2 XENLA STANDARD; PRT: 452 AA.
AC P23770;
DT 01-NOV-1991 (Ref. 20, Created)
DT 01-NOV-1991 (Ref. 20, Last sequence update)
DT 16-OCT-2001 (Ref. 40, Last annotation update)
DB Transcription factor xGATA-2 (GATA binding factor-2).
GN GATA-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Aphiazia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodidae; Xenopus.
NCBI TaxID:5355;
RN 11.
RP SEQUENCE FROM N.A.

```

```

OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
RN RN
RP SEQUENCE FROM N.A.
RS "SSUE=Embryo;
RC MEDLINE=88052679; PubMed=2826129;
RA Ruiz i Altaba A., Perry-O'Keefe H., Meltan D.A.;
RT "Xfin: an embryonic gene encoding a multifingered protein in
RT Xenopus";
RL EMEC J. 6:3365-3370(1987).
RN [2]
RN RN
RP CHARACTERIZATION.
RC MEDLINE=94221366; PubMed=7692399;
RA Andrazzoli V., de Lucchini S., Costa M., Barsacchi G.;
RT "RNA binding properties and evolutionary conservation of the Xenopus
RT multifinger protein Xfin.";
RL Nucleic Acids Res. 21:4218-4225(1993).
RN [3]
RN RN
RP STRUCTURE BY NMR OF FINGER 31.
RC MEDLINE=89346749; PubMed=2501371;
RA Lee M.S., Gippert G.P., Soman K.V., Case D.A., Wright P.E.;
RT "Three-dimensional solution structure of a single zinc finger DNA-
RT binding domain.";
RL Science 245:635-637(1989).
RN [4]
RN RN
RP STRUCTURE BY NMR OF A FINGER.
RC MEDLINE=89378224; PubMed=2506074;
RA Lee M.S., Cavanagh J., Wright P.E.;
RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc
RT finger from Xfin. Sequential resonance assignments and secondary
RT structure.";
RL FEBS Lett. 254:159-164(1989).
CC -1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
CC REGULATION PROCESSES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: COCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC AS NEURAL RETINA CONES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC EMBRYOGENESIS.
CC -1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

```

DR	SMART: SM03349; K2AB; 1.	
DR	SMART: SM03355; ZNF_C2H2; 35.	
DR	PROSITE; PS00805; K2AB; 1.	
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.	
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 37.	
KK	Zinc-finger; Metal-binding; Repeat; 3D-structure;	
KK	Phosphorylation.	
FT	DOMAIN	56
FT	ZN_FING	108
FT	ZN_FING	136
FT	ZN_FING	146
FT	ZN_FING	192
FT	ZN_FING	232
FT	ZN_FING	242
FT	ZN_FING	252
FT	ZN_FING	262
FT	ZN_FING	272
FT	ZN_FING	282
FT	ZN_FING	292
FT	ZN_FING	302
FT	ZN_FING	312
FT	ZN_FING	322
FT	ZN_FING	332
FT	ZN_FING	342
FT	ZN_FING	352
FT	ZN_FING	362
FT	ZN_FING	372
FT	ZN_FING	382
FT	ZN_FING	392
FT	ZN_FING	402
FT	ZN_FING	412
FT	ZN_FING	422
FT	ZN_FING	432
FT	ZN_FING	442
FT	ZN_FING	452
FT	ZN_FING	462
FT	ZN_FING	472
FT	ZN_FING	482
FT	ZN_FING	492
FT	ZN_FING	502
FT	ZN_FING	512
FT	ZN_FING	522
FT	ZN_FING	532
FT	ZN_FING	542
FT	ZN_FING	552
FT	ZN_FING	562
FT	ZN_FING	572
FT	ZN_FING	582
FT	ZN_FING	592
FT	ZN_FING	602
FT	ZN_FING	612
FT	ZN_FING	622
FT	ZN_FING	632
FT	ZN_FING	642
FT	ZN_FING	652
FT	ZN_FING	662
FT	ZN_FING	672
FT	ZN_FING	682
FT	ZN_FING	692
FT	ZN_FING	702
FT	ZN_FING	712
FT	ZN_FING	722
FT	ZN_FING	732
FT	ZN_FING	742
FT	ZN_FING	752
FT	ZN_FING	762
FT	ZN_FING	772
FT	ZN_FING	782
FT	ZN_FING	792
FT	ZN_FING	802
FT	ZN_FING	812
FT	ZN_FING	822
FT	ZN_FING	832
FT	ZN_FING	842
FT	ZN_FING	852
FT	ZN_FING	862
FT	ZN_FING	872
FT	ZN_FING	882
FT	ZN_FING	892
FT	ZN_FING	902
FT	ZN_FING	912
FT	ZN_FING	922
FT	ZN_FING	932
FT	ZN_FING	942
FT	ZN_FING	952
FT	ZN_FING	962
FT	ZN_FING	972
FT	ZN_FING	982
FT	ZN_FING	992
FT	ZN_FING	1002
FT	ZN_FING	1012
FT	ZN_FING	1022
FT	ZN_FING	1032
FT	ZN_FING	1042
FT	ZN_FING	1052
FT	ZN_FING	1062
FT	ZN_FING	1072
FT	ZN_FING	1082
FT	ZN_FING	1092
FT	ZN_FING	1102
FT	ZN_FING	1112
FT	ZN_FING	1122
FT	ZN_FING	1132
FT	ZN_FING	1142
FT	ZN_FING	1152
FT	ZN_FING	1162
FT	ZN_FING	1172
FT	ZN_FING	1182
FT	ZN_FING	1192
FT	ZN_FING	1202
FT	ZN_FING	1212
FT	ZN_FING	1222
FT	ZN_FING	1232
FT	ZN_FING	1242
FT	ZN_FING	1252
FT	ZN_FING	1262
FT	ZN_FING	1272
FT	ZN_FING	1282
FT	ZN_FING	1292
FT	ZN_FING	1302
FT	ZN_FING	1312
FT	ZN_FING	1322
FT	ZN_FING	1332
FT	ZN_FING	1342
FT	ZN_FING	

```

FT ZN_FING 248 270 C2H2-TYPE.
FT ZN_FING 276 298 C2H2-TYPE.
FT ZN_FING 326 348 C2H2-TYPE.
FT ZN_FING 354 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
FT ZN_FING 438 460 C2H2-TYPE.
FT ZN_FING 466 488 C2H2-TYPE.
FT ZN_FING 503 525 C2H2-TYPE.
FT ZN_FING 531 553 C2H2-TYPE.
FT ZN_FING 559 581 C2H2-TYPE.
FT ZN_FING 587 609 C2H2-TYPE.
FT ZN_FING 615 637 C2H2-TYPE.
FT ZN_FING 643 665 C2H2-TYPE.
FT ZN_FING 671 693 C2H2-TYPE.
FT ZN_FING 699 721 C2H2-TYPE.
FT ZN_FING 750 772 C2H2-TYPE.
FT ZN_FING 778 800 C2H2-TYPE.
FT ZN_FING 806 828 C2H2-TYPE.
FT ZN_FING 834 856 C2H2-TYPE.
FT ZN_FING 862 884 C2H2-TYPE.
FT ZN_FING 890 912 C2H2-TYPE.
FT ZN_FING 918 940 C2H2-TYPE.
FT ZN_FING 988 1010 C2H2-TYPE.
FT ZN_FING 1016 1038 C2H2-TYPE.
FT ZN_FING 1044 1066 C2H2-TYPE.
FT ZN_FING 1136 1158 C2H2-TYPE.
FT ZN_FING 1164 1186 C2H2-TYPE.
FT ZN_FING 1192 1214 C2H2-TYPE.
FT ZN_FING 1220 1242 C2H2-TYPE.
FT ZN_FING 1248 1270 C2H2-TYPE.
FT ZN_FING 1276 1298 C2H2-TYPE.
FT STRAND 1045 1045
FT STRAND 1052 1052
FT HELIX 1056 1062
FT HELIX 1063 1065
FT HELIX 1066 1066
SQ SEQUENCE 1350 AA; 27F10AB0851E0AD8 CRC64;

Query Match
Best Local Similarity 3.5%; Score 238; DB 1; Length 1350;
Matches 239; Conservative 137; Mismatches 390; Indels 476; Gaps 70;

QY 222 PKNICGGYGNPTDLIKHFRVHGLHNRTRQDAELDSKILALHNMVQF----- 273
DB 354 YLCSHCNKGFIQ--SDLVKHFT-HTG--ERPQCAECHKGFIQKSDLVKHLATH-GEK 408
QY 274 ---SHSKDFOKVNRVSFGLVDINSSRPVLLNGLTYDVQVTSGGTF:GIGRKTPTDC--- 326
DB 409 PFKSHC-DKXFTERSALAKHQRTHTGKPY-----KSCDCKE 446
QY 327 -----QNTKYFRCKFCNFTYMGNSSTLEQHFLOTHPNKIKASLPSEVAK 373
DB 447 FTQRNLILHQRHTGERPYKTLCDRTFIQNS--DLVKH-----QKVHANPLSD--- 495
QY 374 PSEKNS-----NKSIPALOS-----SDGDLGKQWDKITV 403
DB 496 PHTANSPHKSKCDLTFGHNSTFMKSHLHSGEKFKQCAECKGFTQKSDLVK---HIRV 552
QY 404 KAGDDTPGVSVPIKPL---DSSRQNGTEATSYW-----CKFCSESSSSLK-LJ 452
DB 553 HTGEK-----PFKLLCKKSFSQNSD--LHKHWRHTGKPPCYCTDKSFTERSALI 603
QY 453 EHYGQHGAVOSGGLNPELNDKLRSVINQNDLAKSSEGTMTKTDKSSSGAKKDFSS 512
DB 604 KHRHTHTG-----ERPHKCSVCQKGFQKSA---LKHSEHTHGEK----- 642
QY 513 XGAEDNMVTSYQCDFRYSKSHGPDVIVGVLRLHVQQLHNTHK---CT----- 560
DB 642 -----PYPTQC-----GKSPIQNSDLVKH-QR-HTGKPDVHCTECKKFTTEG 683
QY 561 ---LKH-----CPFCRGLCSPE---KH-----GE-----ITYPFACRKSN-- 591

```

```

DB 684 SSLVXKRRTHSGEKPYRCPCQCEKTFIQSSDLVKHLVWANGENPPAATAFHEILIRRENLT 743
QY 592 -----CSHCA-----LALLHLSPGAGSSRVKQCHCCS---FTTPDQVDVLFHY 633
DB 744 RSEPDYPCTCEGKVPHGCRPALLKHL---RTHKTEKYPCHNECDKSPQTSD-----LVKH 796
QY 634 ESHVESQASDVKQEAN--HLGSGSQSQSVKESKEH-----SCTKCD--FITQVEEELSR 683
DB 792 LRTH-GERPYHCPENCKGFIQNSD--LVKHQPTHGERPYTCSCODKGFQIQR--SALTU 851
QY 684 HYR--RANSCYKCRQCSFTADTQSLLEHFNTHVHCQEQDITTANGSEEDCHAISTKEEPK 742
DB 852 HNRHTHTGKPYKCECCQKCFQNSDLVKH-QR-H-----TGEKPYHC----- 892
QY 742 IDPRVYNLTTPSKMGEPPVSESVKREKLEK-----DGLKEKVMTESSSDLRNVTWK 796
DB 893 -----PCKKRFTEGSSLLKHRIHSRIKPYPCGVCKGKSFQSS----- 931
QY 797 ADLRGSPSYTOASLGLTPVSGTQE-----QTKLRDSNPVEAAHLARPIYGLAVETKG 851
DB 932 -NLKELKXCHSEON-----PPVALSSELGFAVETQTHPDVQ-----H1---VVG---DTAS 976
QY 852 FLOGAPAGEKSGSALPOQYPASGENKSKDECSLLRRRAGSGVFCANCLTTKSLWRKNA 911
DB 977 YI-SPEAAGERS-----LYOKLHSTPRPLNIIKONNGEOIRRRTRKLNPEALCAE 962
QY 988 ---FKNCCGKCFARHSV---KXVR-HTGERPYK-----SOCTRSFIQXS 1029
QY 962 QLNKQORNSSEVNGSFLERS--EDHLTESHQREIPLPSLSKYEAGSGTJKSHSAQOP 1029
DB 1030 DLVKHYR-HTGERPYKGLCERSFVEKSGALRRHQ-----VHKNESP 1071
QY 1020 VLVSQTLDIHKEMOPLHLCIKSPQESTGDPGNSSSVSGKSSBEGSP--LEKVMRPAKH 1077
DB 1072 VL-----NSAME--QCCVTYMGSKDDP--NSLVQPLHVIKEESPHIVNAYSPLSIL 1123
QY 1076 PNYSP-----GSPIEKYQYPLFGLPFVHNDQSEADMLRFWSKYKLSVPGN----- 1124
DB 1121 QSYFPILPEPKGP-RYSCSECGKCFHRSV-----FLKHWRYHTGEOPTCKECCGSF 1173
QY 1125 -----PHYLSHVGLPNPCQNYVPYPTFNLPHFSAVGSND:PLDLA:KHSPGPTA 1177
DB 1174 SSSALVKVRIHTGKPYC-----STCGKSFIOKSDLA-KHQR----- 1213
QY 1178 NGASREKTKAPENVKNEGPLNVVTKYKVDKDRSTQDELSTKYHCGIVF:DEVYALHMSCH 1237
DB 1214 HTGKPYTCVCGKFPDRSSVVKH---SRHTGERPYKNECTKGFVQKSDLVKHMRTH 1270
QY 1238 GDSGPFQCSICCHLCTDKYDFTTHIQRLHRNNAQVEXNKP 1279
DB 1271 TGEKPYGONCCOR-----SFSTHSASVRHQ---RMCNTGRP 1303

RESULT 14
AREA GISFC
AC AREA GISFCU STANDARD; PRT: 971 AA.
ID P7868;
CT 15-JUL-1998 (Rel. 36, Created)
CT 15-JUL-1998 (Rel. 36, Last sequence update)
CT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus; Fusarium moniliforme).
CC Eukaryota; Fungi; Ascomycota; Pezizomycotina, Sordariomycetes;
OX Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
RN NCBI TaxID=5127;
RP SEQUENCE FROM N.A.
RC STRAIN=ms67;
RX MEDLINE=59:69774; PubMed=10071216;

```

RA Tudzynski B., Homann V., Peng B., Marzluf G.A.:
RT "Isolation, characterization and disruption of the area nitrogen
RL regulatory gene of *Gibberella fujikuroi*.";
RL Mol. Gen. Genet. 261:106-114(1999).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Y11006; CAA71897.1; --
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN FING 694 718 GATA-TYPE.
SQ SEQUENCE 971 AA; 103580 MW; 887DD88214C7453 CRC64;

Query Match 3.4%; Score 234.5; DB 1; Length 971;
Best Local Similarity 21.0%; Pred. No. 8.7e-06;
Matches 181; Conservative 98; Mismatches 313; Indels 269; Gaps 41;

QY 344 GNSSTLEQHFLOTHPNKIKASLPSE--VAKFSEKN-----SNKSIPLAQSDSGDL 394
DB 168 GSAFRLGRPMQSSFGIAQLAKSSENNLAQDAVNLDGFIQSDNGSGPINFASPEGDX 227
QY 395 -----GKMQDKITVAGSDTP-----VGYSVPIKPLDSRRNGTEATSYVCKFCPSFC 443
DB 226 MVDRRSGSSMASAIPKSRKEPSLQNFVQSPVQFAHQATQ-GSE-FYVNVNRH-RKTSI 285
QY 444 ESSSSIKLLEHYCKHCAVQSGNLDELNDKLSRGVINGNDLAKSSEGETYKTKCKSS 503
DB 286 DDRTRKRPANFSPQPAV-----NSTAQNDELDELSELDHVSJLQCPNGA 330
QY 504 GAKKDFSSKGA-----EDNMTSYNQCFDFRYKSHGPDVIVWGLLPHYQQ 552
DB 331 GIPQ---SNGSNVPFEMIDTFEMENDSMWNGNFQ-QNFSFSPSP-MIPGHFSGMY-- 383
QY 553 LNIHKTIKHCFPCPGLCSPEKHGEITYPACRKNCSHCALLLLHLSPGAGSSRY 612
DB 384 -HN-----SSVPSASMSNNNSD-----FSPASAYPSNV 414
QY 613 KHQCHQCSFTPDVDVLLPHVESVHESQASDVQKQEAHNL-----QGSQGOQ-SVKESK 664
DB 415 -----STPHVPVQEGFYFGSCQARTCPQGFQCSITGSMLSQGFMYGNGSGSTYFSA 469
QY 665 EHCTKCDFTIQVEEISRRYRAHSCYKCRQCSFTAAADTQSILLEHFNTHVHCOEDITTA 724
DB 470 PGTASE-----SMSAYSTAPSSFGHIDPSQVFQ-----EQAVTS- 504
QY 725 NGEDGHAISTIKKEPKIDFRVNVNLLTPDSKMEPVSESVVREKLEEKGLKKEKWTES 784
DB 505 -----PTIQPQNMFSFGADSDDEDNNAFADRVSMQKD-----MSS 542
QY 785 SSDLRNVTVRG-----ADLRGSPSYTCASLGILLTPVSGT----- 820
DB 543 SLDESGANGDASLPGCFSTQAAAFPGGPRKQWNG-----GTTDTFDVNDGNGWESNG 596
QY 821 --QBQTKLRDSPNVEAH-----LAPTYGLAVETKGF-----LQAPAG---GEKS 963
DB 597 LERSQSQSRFGG-NLRQHPKLPENASTPVH-FGGQNGFEQLAQSMQSPAGDNGTWS 654

RESULT 15

IKAR_VQJUSE STANDARD; PR7; 517 AA.
AC Q03267; Q64044; C64045; Q64051;
ET 01-OCT-1993 (Re: 27, Created)
ET 15-DEC-1998 (Re: 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein Ikaros (lymphoid transcription factor Lymph-1).
GN ZNFN1A1 OR IKAROS OR LYF1.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
PN 11
RP SEQUENCE FROM N.A. (ISOFORM V1).
RC TSSJ2=Embryo;
RX MEDLINE=91368267; Pubmed=1439790;
RA Georgopoulos K., Moore D.J., Deifler B.;
RT "Ikaros, an early lymphoid-specific transcription factor and a
RT putative mediator for T cell commitment.";
RL Science 258:908-912(1992).
RN 12
PP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RA MEDLINE=95021239; Pubmed=7935426;
RA Hart K., Ernst P., Lo K., Kim G.S., Turck C., Smale S.T.;
RT "The lymphoid transcription factor Lymph-1 is encoded by specific,
RT alternatively spliced mRNAs derived from the Ikaros gene.";
RL Mol. Cell. Biol. 14:7111-7123(1994).
CC -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC CONTROL ELEMENT IN THE TCT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED
CC DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoform=s6;
CC Name=V1;
CC IsoId=Q03267-1; Sequence=Displayed;
CC Name=V2;
CC IsoId=Q03267-2; Sequence=VSP_006855;
CC Name=V3;
CC IsoId=Q03267-3; Sequence=VSP_006853, VSP_006855;
CC Name=V4;
CC IsoId=Q03267-4; Sequence=VSP_006856;
CC Name=V5;
CC IsoId=Q03267-5; Sequence=VSP_006853, VSP_006856;
CC Name=V6;
CC IsoId=Q03267-6; Sequence=VSP_006854;
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN T-CELLS AND THEIR
CC PROGENITORS, AND ALSO IN B-CELLS.

CC -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: L03547; AAB66193.1; -
DR EMBL: S74517; AAB32248.2; ALT_SEQ.
DR EMBL: S74518; AAB32249.2; -
DR EMBL: S74708; AAB32250.2; -
DR HSP: P15822; IBB0.
DR TRANSFAC: T01470; -
DR MGD: MGI:1342540; Znf141.
DR GO: GO:0003677; F:DNA binding activity; IDA.
DR GO: GO:0003097; P:hemopoiesis; IMP.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IMP.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR ProDom: PD000003; Znf C2H2; 2.
DR SMART: SM00355; Znf C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT ZN_FING 117 139
FT C2H2-TYPE 1.
FT ZN_FING 144 166
FT C2H2-TYPE 2.
FT ZN_FING 172 194
FT C2H2-TYPE 3.
FT ZN_FING 200 223
FT C2H2-TYPE 4.
FT ZN_FING 457 479
FT C2H2-TYPE 5.
FT ZN_FING 488 512
FT C2H2-TYPE 6.
FT VARSPPLIC 53 53
M -> VAYGAGGDFRDFHALISDRGM (in isoform I)
and isoform IV).
FT VARSPPLIC 54 140
FT VARSPPLIC 54 140
FT VARSPPLIC 54 282
FT VARSPPLIC 141 282
FT VARSPPLIC 141 282
FT CONFLICT 234 235
FT CONFLICT 480 482
FT CONFLICT 517 AA; 57336 MM; 1052B8E76AF24287 CRC64;
SQ
Query Match 3.4%; Score 234; DB 1; Length 517;
Best Local Similarity 23.6%; Pred. No. 3.9e-06;
Matches 127; Conservative 63; Mismatches 204; Indels 144; Gaps 25;
QY 814 LTFVSGTQQTTLRD-SNVFAAHLARIYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
DB 37 LSTTGAQGNKSDRGMSNVK-----VETQSDENGRACENNGEACEDLRM 84
QY 870 YPASGENKS---KDESQSLRRRG-----SQVFC--ANCLTKTSLWRKNAN 912
DB 85 LDASGEKMGSHRDQGSALSQVGGIRLPNGKLCDCIGVICGNVLMVHK---RSHT 141
QY 913 GGVVNCAGJ-----YQKLHSTPRPLNIKKNGEQIIRRTKRRLNPEALQAEQ 962
DB 142 RPFQNCOCASFTQXGNLLRHKLHSGKEP---FKCHLQYACRR-----DALTCHL 191
QY 963 LNKQQRGSNEEQVNGSPLRRS--EDHLTESHQ--REIPJPSL-----SKYEAQCSL 201C
DB 192 RTHSVGKPKHKGCGYGRSYKQRSLEBHKRCHNYLESMLPGVCPVKEETNHNEMAEGL 251
QY 1011 TKSHSAQQPVLYSQTLDIHKWQPLHIQIKSQESTGDPGNSSSVSEKSGSSEKSPLEK 207C
DB 252 CKIGAERSLVLRASNAVARKSM-----PQKFLGDKCLSDXPYDSANYEKEDMTSH 305
QY 1071 YMRPAKHP--NY-----SPFGS-----PIEKYQY-----PLFGLPFVHNDQSEA 1128

DB 306 VMDQAINNAINYLGAESLRPLVQTPPGSSSEYVFISSMYQLHKPPSDGPPSRNSACDAV 365
QY 1109 DWLREWSKYKLSVPCNPHYLSGHVPGLP-NPCQNVVP-----YPTFNLPEHFS 1154
DB 366 DNJLLSKAK-SVS-----SEREASPSNSQSDSTDIESNAEQRSGLIYLTNHNPH-- 416
QY 1155 AVGSNDIPLDLA:KHSRPGPTANGASKETKAPFNKNEGPJNVVTEKYDRSTQDELS 214
DB 417 -----ARNGALKERCAVEVLRAASENSQDAFRVVSSTGEOLK 455
QY 1215 T-KCVHCGIVFELEFNVYALHMSCHDSQ---PFGCSLCCHLCTCKYDPTTHIGRLHR 1268
DB 456 VYKCEHCPLVLFJLHVMYTHMGCHGCHGFRDFFECNNGYSGYSDRYEFSKTRGEER 513
Search completed: October 29, 2003, 11:30:53
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:57 ; Search time 191 Seconds
(without alignments)
1730.709 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKKNPLRNVAEGEQI.....TQRGJHRNNAQVKNQKPKX 128:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertibrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5257.5	76.7	1092	11 Q81L70	Q81L70 mus muscu
2	923	13.5	229	11 Q82B62	Q82B62 mus muscu
3	282	4.1	1852	11 Q8CHH6	Q8CHH6 mus muscu
4	271	4.0	1186	11 Q89G61	Q89G61 rattus norv
5	267.5	3.9	1224	4 Q9NZ13	Q9NZ13 homo sapien
6	265.5	3.9	1311	11 Q8CIQ2	Q8CIQ2 mus muscu
7	265	3.9	1271	11 Q8CIQ1	Q8CIQ1 mus muscu
8	264.5	3.9	1311	11 Q8VDS6	Q8VDS6 mus muscu
9	263.5	3.8	1669	11 Q925J5	Q925J5 mus muscu
10	262.5	3.8	589	13 Q8UVK1	Q8UVK1 petromyzon
11	262	3.8	1167	11 Q9ESD2	Q9ESD2 mus muscu
12	261.5	3.8	1669	11 Q925J4	Q925J4 mus muscu
13	260	3.8	537	13 Q93581	Q93581 brachydanio
14	258	3.8	507	13 Q918U5	Q918U5 petromyzon
15	256	3.7	1311	4 Q96K83	Q96K83 homo sapien
16	256	3.7	1412	4 Q96JM2	Q96JM2 homo sapien

ALIGNMENTS

RESULT :

Q8K1C0

ID Q8K1C0

AC Q8K1C0

DT C1-CCT-2002 (EMBLrel. 22, Created)

DT C1-CCT-2002 (EMBLrel. 22, Last sequence update)

DT C1-VAS-2003 (EMBLrel. 21, Last annotation update)

DE Similar to trichorhinophalangeal syndrome 1 (Human).

GN TRPS1.

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC037058; AAH37358.1; -

DR M3C; MG1:1927616; Trps.

DR InterPro; IPR007087; Znf_C2H2.

DR InterPro; IPR006675; Znf_GATA.

DR Pfam; PF03320; GATA_1.

DR PRINTS; PR00619; GATAZFINGER.

DR SMART; SM00355; Znf_C2H2; 6.

DR SMART; SM00401; Znf_GATA; 1.

DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.

DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.

DR PROSITE; PS00328; ZINC_FINGER_C2H2_1; 2.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.

DR Metal-binding; Zinc; Zinc-finger.

SC SEQUENCE 1092 AA; 120053 MW; 216BC4B8C91C89EE CRC64;

Query Match: 76.7%; Score 5257.5; DB 11; Length 1092;

Best local similarity 77.7%; Pred. No. 3;

Matches 1034; Conservative 35; Mismatches 42; Indels 211; Gaps 2;

Qy 1 MVRKKNPLRNVAEGEQI...LEPTIGTSKVSQGNKKEFSADQKSENTOQSCAKELNHKEH 50

Db 1 MVRKKNPLRNVAEGEQI...LEPTIGTSKVSQGNKKEFSADQKSENTOQSCAKELNHKEH 60

Qy 1 MVRKKNPLRNVAEGEQI...LEPTIGTSKVSQGNKKEFSADQKSENTOQSCAKELNHKEH 50

Db 1 MVRKKNPLRNVAEGEQI...LEPTIGTSKVSQGNKKEFSADQKSENTOQSCAKELNHKEH 60

```
QY 61 SLHVQDPSSSSKKDLKSAVLSEKAGFNYESPSKGNFSPFFPHDEVTDNRMLAFSPFAAG 120
DB 61 STHGQEPSSSGKKDLQISGLSEKAGFNYESPSKGSLSVFFPHDEVTDNRMLAFSPFAAG 120
QY 121 VCEPLKSPQRAADDPQMACTPSSGDSLETEDQKMSPKATEETGQASQGANCGGLSPV 180
DB 121 VCEPLKSPQRAADDPQMACTPSSGDSLETEDQKMSPKATEETGQASQGANCGGLSPV 180
QY 181 SVASKNPQVPSDGGVRLNKSITDLLVNPNPPAPLSPRELQGFKNICGYGYGNPDTCLI 240
DB 181 SVASKNPQVPSDGGVRLNKSITDLLVNPNPPAPLSPRELQGFKNICGYGYGNPDTCLI 240
QY 241 KHFYKYLHLGNHRTQDAELDSKILAJLHNMVQFSHSKDFQKVRNSVPSGVLDQ:NSSRPV 300
DB 241 KHFYKYLHLGNHRTQDAELDSKILAJLHNMVQFSHSKDFQKVRNSVPSGVLDQ:SSSRFA 300
QY 301 LINGTYDVQVTSGGTFIGRKTPOCGNTKYPRCKFCNFTYMGNSSTELBOHFLQTHPN 360
DB 301 LINGTYDVQVTSGGTFIGRKTPOCGNTKYPRCKFCNFTYMGNSSTELBOHFLQTHPN 360
QY 361 KIKASLPSEVAKPSEKSNKSIKIPALQSSDSGLCKWQOKTVKAGDDTPVGYGVPIKPL 420
DB 361 KIKASLPSEVAKPSEKSNKSIKIPALQSSDSGLCKWQOKTVKAGDDTPVGYGVPIKPL 420
QY 421 DSSRQNGTEATSYWCKFCSPSESSSSKULEHYGKHGAVQSGGLNPELNDKLSKGV 480
DB 421 DSSRQNGTEATSYWCKFCSPSESSSSKULEHYGKHGAVQSGGLNPELNDKLSKGV 480
QY 481 INQNDLAKSSSEGETWTKTKDKSSGAKKXDFSSKGAEDNMVTSYKCCPCDFRYSKSHQPDV 540
DB 481 INQNDLAKSSSEGETWTKTKDKSSGAKKXDFSSKGAEDNMVTSYKCCPCDFRYSKSHQPDV 540
QY 541 YVGPLRLHYOOLHNIHKTIKHCPFCPRGLCSPEKHILGE:TYFPACKRKNCSHCALJLL 600
DB 541 YVGPLRLHYOOLHNIHKTIKHCPFCPRGLCSPEKHILGE:TYFPACKRKNCSHCALJLL 600
QY 601 HUSPGAASSRKYKHCHQCSTFTPDVYVLLFHYESVHESQASDVQKQENHLOGSGGQSV 660
DB 601 HUSPGAASSRKYKHCHQCSTFTPDVYVLLFHYESVHESQASDVQKQENHLOGSGGQSV 660
QY 661 KESKHSCTKCDPITQVEEISRHVYRAHSCYKQCQCSFTAADTQSLEHENTVHCQSD 720
DB 661 KESKHSCTKCDPITQVEEISRHVYRAHSCYKQCQCSFTAADTQSLEHENTVHCQSD 720
QY 721 ITTANGEEGHAISTIKPEPKIDFRVYLLTPDSKMGEPVSESVYKREKLEKDGKLEK 780
DB 721 ITTANGEEGHAISTIKPEPKIDFRVYLLTPDSKMGEPVSESVYKREKLEKDGKLEK 780
QY 781 WTESSSDDLRYNTRGADILRGSPSYTOASLGLLTPVSGTOETKTLRDSPNVEAAHLAR 840
DB 781 WTESSSDDLRYNTRGADILRGSPSYTOASLGLLTPVSGTOETKTLRDSPNVEAAHLAR 840
QY 841 PIYGLAVETKGLQAPAGKSGALPQCPASGENKSKDESQSLL:-----RRR 899
DB 841 PIYGLAVETKGLQAPAGKSGALPQCPASGENKSKDESQSLL:-----RRR 899
QY 890 RSGGVPCANCLTTKTSLRKKNANGVYVNCAGLYOKLHSTPRPLNIKQNGEOIIRRT 949
DB 890 RSGGVPCANCLTTKTSLRKKNANGVYVNCAGLYOKLHSTPRPLNIKQNGEOIIRRT 949
QY 950 RURLNPEALQABLKNQKQGSNEEONGVNSPLERRSECHLTESHOREIPLPLSLKYEACGS 1009
DB 950 RURLNPEALQABLKNQKQGSNEEONGVNSPLERRSECHLTESHOREIPLPLSLKYEACGS 1009
QY 1010 LTKSHSAQQVPLVSQALDIHKRMQPHIQIKSPQESTGDPGNSSSVSEKSSSERGSPIS 1069
DB 1010 LTKSHSAQQVPLVSQALDIHKRMQPHIQIKSPQESTGDPGNSSSVSEKSSSERGSPIS 1069
QY 1070 KYMRPAKHPNYSPGSPIEKYQYPLFGPLFVHNDQSEADWLRFMSKYKLSVPQKNHYLS 1129
DB 1070 KYMRPAKHPNYSPGSPIEKYQYPLFGPLFVHNDQSEADWLRFMSKYKLSVPQKNHYLS 1129
QY 881 KYMRPAKHPNYSPGSPIEKYQYPLFGPLFVHNDQSEADWLRFMSKYKLSVPQKNHYLS 940
DB 881 KYMRPAKHPNYSPGSPIEKYQYPLFGPLFVHNDQSEADWLRFMSKYKLSVPQKNHYLS 940
```

```
QY 1130 HYPGLPNPCQNVVYPTFNLPHPFSAVGSNDIPLDLAIKHSRPOPTANGASKETKAPP 1189
DB 941 HYPGLPNPCQNVVYPTFNLPHPFSAVGSNDIPLDLAIKHSRPOPTANGASKETKAPP 1000
QY 1190 NYKNGEPLNVYKTEKVDKDRSTQDELSTKCVHCGIVFDEMYALHNSCHGDSGPPCCSICQ 1249
DB 1001 TVKNEGEPNWWKTEKVDKDRSTQDELSTKCVHCGIVFDEMYALHNSCHGDSGPPCCSICQ 1060
QY 1250 HCTDKYDFTTHIRGLHNNNAQVEKNGKPK 1281
DB 1061 HCTDKYDFTTHIRGLHNNNAQVEKNGKPK 1092
RESULT 2
Q8Z52 PRELIMINARY: PRT: 229 AA.
AC Q8Z52;
DT 01-VAR-2003 (TRENDELrel. 23, Created)
DT 01-VAR-2003 (TRENDELrel. 23, Last sequence update)
DT 01-VAR-2003 (TRENDELrel. 23, Last annotation update)
DE Trichorhinophthalangeal syndrome 1.
CS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN 1.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; Tissue=Hore;
RX MEDLINE=23354683; PubMed=12466851;
RA The FANTOM Consortium;
RT The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
CR BMSL; AK036590; BAC29495.1;
SQ SEQUENCE 229 AA; 25290 MW; 03283E1B7AABA039 CRC64;
Query Watch: 13.5%; Score 923; DB 11; Length 229;
Best Local Similarity 85.6%; Pred. No. 2,3e-54;
Matches 173; Conservative 17; Mismatches 12; Indels 0; Gaps 0;
QY 695 RRAHSCYKQCQCSFTAADTQSLEHENTVHCQSDITTANGEEGHA:STIKESPKIDFR 745
DB 20 RRAHSCYKQCQCSFTAADTQSLEHENTVHCQSDITTANGEEGHA:PTIKESPKIDLK 79
QY 746 VYNLLTPDSKMGEPVSESVYKREKLEKDGKLEKDKKWTESSDQDLPNVTWAGADILRGSPS 805
DB 90 VYSLNPDSSKMGETVPES:VKREKLDKKEGDKDKWTESSTDDLRGVAVRGADILRGSPS 139
QY 506 YTOASLGLLTPVSGTOETKTLRDSPNVEAAHLARPIYGLAVETKGLQAPAGKSGA 865
DB 140 YTOASLGLLTPVSSQEQTKTLRDSPNVEAAHLARPIYGLAVETKGLQAPAGSEKAS 199
QY 666 LPOQYPASGENKSKDESQSLLR 887
DB 200 LPOQYPASGENKSKDESQSLLR 221
RESULT 3
Q8CHH6 PRELIMINARY: PRT: 1852 AA.
AC Q8CHH6;
DT 01-VAR-2003 (TRENDELrel. 23, Created)
DT 01-VAR-2003 (TRENDELrel. 23, Last sequence update)
DT 01-VAR-2003 (TRENDELrel. 23, Last annotation update)
DE KIAA0236 protein (Fragment).
GN KIAA0236.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN 1.
RP SEQUENCE FROM N.A.
```

DB	852	EHEIVANSNMOALPETHNEEASPKRQ----	KGIE-----	AFQEDQVD-----	890						
QY	803	SPSYTQASLGJLTPVSGTCEQTKLRD----	SPNVEAAHARFIYGLAVETKGFLOGAPAGG	56C							
CB	891	SSSLGEVREG-----	GCTLHLEALRVEPTEPEPLPLEELTETATV	944							
QY	861	EXSGALPQOYPASGENKSKDEQSILLRRRGSGVFCANCLTTKTSLRKRNKAGGYVCNAC	920								
CB	945	ERPGLER--PALSEPDS-IETPALVAEE-----	971								
QY	921	GLYQKLHSTPRPLNIKKNNGSQITARRTRKRLN-PEALCAEQLNKQCSNEEQVNGSP	979								
DB	972	PVVEK-LASEPPRPLISEBAPNTEKAAALTAETVPJPPFPESESLKAMSEQDKEQAEALV	1031								
QY	960	LERRSE-----	CHLTESHQRIPLPS-SKYBAQGSGLTKSHSAQOPVLVSQTL	1026							
DB	1032	LEGRVQVVVIQEGRAFRCPHCPTRTRKALTLSKSCQG-----	RRPPLCFEGC	1084							
QY	1027	DHKRMQPL-HQIKSPQESTDPCGNSSSVSEGGKSSERGSPIEKYMPAKHP-----	1078								
DB	1085	ASFKQQRGLSTHYMKKCPV-----	LKKNAALPKFVSFTLHPQ-PCNQ	1127							
QY	1079	-----	NYSPSPSPIEKYCYPLGLEFFVNDQSEADKILRFNSKYKLSVPGN-PIYLSHV	1131							
DB	1128	ASQDAESRXPPPL-----	SKVELLLPKDAPSLPGS	1159							
QY	1132	PGLNPQCNVVPYPTNLPFHFSAVGSDKIDPLDLAKHSRPGFTANGASKETKAPPNV	1191								
DB	1160	FGVEELPTFSDFPT--SPD-----	ENSLP-----	GTSEK-----	1188						
QY	1192	KNEGPNVYKTEKVDIRSTQDELSTKCVHCGIVFLDEVMYALHWYSGDSGPPQCS-CQHL	1251								
DB	1189	-----	PHFEQGKPHCSSCTFL	1204							
QY	1252	CTDKYDFTTHIQGLHRRNAQVEKQKPK	128C								
DB	1205	CARLSSITSHVTGGCGRGGRQKMGGRQ	1233								
RESULT 4											
COB#61											
ID	COB961	PRELIMINARY;		PRT;	1:86 AA.						
AC	COB961;										
CT	01-JUN-1997	(TEXTBLRE: 04, Created)									
CT	01-JUN-1997	(TEXTBLRE: 04, Last sequence update)									
CT	01-MAR-2003	(TEXTBLRE: 23, Last annotation update)									
RE	Reaz.										
CS	Rattus norvegicus (rat).										
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.										
OX	NCBI_TaxID=10116;										
RN	[1]										
SEQUENCE FROM N.A.											
RF	STRAN=Sprague-Dawley;										
RA	Tsai R.Y.H., Reed R.R.;										
RT	Cloning and Functional Characterization of Reaz, a Zinc Finger Protein										
RT	that Interacts with C/EB β to Regulate Gene Expression: Implications										
RT	for Olfactory Neuronal Development.;										
RL	Submitted (MAR1997) to the EMBL/GenBank/DCBJ databases.										
DB	EMBL; J92564; AAB58646.1;										
CR	HSSP; P08047; 1SP.										
CR	TRANSFAC; T02815;										
DR	InterPro; IPR007087; Znf_C2H2.										
DR	Pfam; PF00036; zf_C2H2_28.										
DR	ProDom; PDC00003; Znf_C2H2_2.										
DR	SMART; SM00355; Znf_C2H2_28.										
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 27.										
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 21.										
KW	Metal-binding; Zinc; Zinc-finger.										
SQ	SEQUENCE 1186 AA; 133479 MW; 562D6C779BCB9FCA CRC64;										
Query Match											
Query 1186; DB 11; Length 1186;											


```
Best Local Similarity 20.0%; Pred. No. 2.4e-09;
Matches 275; Conservative 156; Mismatches 515; Indels 432; Gaps 69;

QY 52 ALNKHKEHSLHVPSPSSSKDLKASAVLSEKAGNYESPSKGNFSPFPDEVDYDNKL 11:
DB 84 AAFSRDLHLKHLKTHSS--KPFKCSVC--KRGFSSTSS--CS-----HMQAHKXNK- 131:
QY 112 AFSFAAGVCEPL--KSPORAEADPPQWACTTPSGDSL--ETKEQK-----KSPKATEETG 165
DB 132 -----EHLAKSEAKKO--DFMCDYCEDTFSQTEELEKHLVLTLEPQLSEKAD 177
QY 156 -----QAQSGQANCOGLSPVSA-----SNXPQ 188
DB 178 LQCHICPEVFDSTLLAHIHOAHQKHCKPCPEQFSVGVYCH--DSHRQPSDSNHS 237
QY 189 VSPD--GGVRJNKSXTDLLVNDNPP-----APLSPELCQFKNICGYGYGNDPTLLI 240
DB 238 VSPDPVLGVSASMSSATP--DSTPDPVJGVSASMSSATPSSASV-----ERGSTPOSTL 290
QY 241 KHRKYHLGLHNRTRQDAELDSKILALHNMVQFQKVRSVFSGVLQCNSSRPV 300
DB 291 KPLRG-----QKKVRDDGGSWKVV-----YSCPYCKSRDFTSL--AVLEIHLKTIHACKP- 339
QY 301 LLANGYDVQVTSGETFIGIG--RKT-----PDQ--QNTXYFRCKFCNFTYXNSSTEL 350
DB 340 QQSHTCCQICLDSMPTLYLNNEHVRKLHKSHPVPMQFNGISAFHCNICYPEFADINS--L 397
QY 351 EGHFLQTH--PNKIKASLPSSSEVAKPSEKNS---NKSIPALQSSDGLGKQCKTVK 404
DB 398 QEHIRVSHCGPN-----ANPPGNAFFCNCQSMGFLTESS--J-EHILQQAACS 444
QY 405 AGD---DTPVGVSPiKPLDSSRQNGTEATSYWCKFCSPCESSSSIKLL-EHYKQKGA 461
DB 445 VGSTKLESVP-----IQPTQSPVE-----VYSCPYNCTNSIPFSGILKTKIKEN:GN 492
QY 462 VQSGGLNELNDKLRGVSINQNDLAKSSEGTMTKTDKSSGAKKDFSKGADNMVT 521
DB 493 -----PLAHSK-----KSKABQSPVSSJVEVSSPKRQLSGSA-----NSIS 530
QY 522 S--YMCQCFDRYSKSHGPDVIWGPLLRHVQQLNIHK--IKHCFFCPGRGLCKPEKLG 579
DB 531 NGEYCNQCDLKFNFES-----FQTHLKLHLEULLRKQACPCQCKEDPDQESILQ 581
QY 580 EITYFPACKSN--CSHC-----ALLHLSPGAAGSRVVKHQCHQCSFTTPD 625
DB 582 HLTVMYMTTSTHYVCESCDKFSSVDLQKHLDWH-----TFVLYHCTLCQEVFDS 633
QY 626 VQVLLFHYESVHESQAS-----DYKQEAHLOGSDGQGVKES-----KHSCTK 670
DB 634 KUSIQVHLAVKSHNEKMYRCTACNWDFFREA-----CLQHVHSHLGNPAKAHKCIF 687
QY 671 CDFITQVEEISRHVYRAHSCYKRCQCSFTAADTQSL-EHFNVTVHCQCDJITFANGEEJG 730
DB 688 CGETTS-EVELOCHITTHSKYKNCRCFCAHFAVLELLEKHLREKHC-VFDPAENGATANG 746
QY 731 HAISIKESPKIDFVYNLLTTPDSKMGEPVSVVYKREKLEKGC:KKNVTESSDZL- 789
DB 747 VPTSTKKAEPADLOMLKNPEA-----PNSH-----BASEDQVD 782
QY 790 RNVTRGADILRGSPYTCASLGLLTTPVSGTCQTKTLRDSPNVEAAHLARPIYSLAVET 849
DB 783 ASEPMVGDIC--GAAYTVEVL-----LQHRLAD-----HNIRP----- 815
QY 850 KGFLOGAPAGGKSGALPOOYPASGENKSKDESQSLRRRRSGGVFCANCLTKTSLMRK 909
DB 816 -----GEDGSRKKAEIFKSHKCNVCSRTF-----FSENGLEHLQTHRG 956
QY 910 NANGGVVNCAGC:YQKLHSTPRPLNI:KQNGEQILIRTRKRLPAPEALQAEQLNKQORG 969
DB 857 PAK-HVMCPICG-----ERFPSLLTLEH-----KVT-HSKSLD 888
QY 970 SNEEQVNGSPLERRSRDLHTESHOREIP:PSLSKYEAQGSLLT--KSHSAQCPVLVSOTLJ 1027
889 TGTCTCRICKMPLQ--SEEBPIEHQCMH---PDL-----RNSLTGPRCVVCMQVTTSTLELK 938
1028 LHK--PMQQLHLQIKSPQSGTCDPGKSSSVSGKSGSRSGSP:EKYMRPAKHPNYSPEGS 1085
939 IGTTFHMQKL-----AGSSAAS-----SPNQ 960
1086 PIEKYCYP:FGLPFPVHNDPQSEADMLRFWSKYKLSVPGNPHYL-----SHVPGI- 1134
961 GLQK-----LYKCALCLKEFRSKQD-LVR-----LVNGLPYGLCAGCAXSANGQVGLA 1010
1135 -NPKQNYVPYPTFFLPPHFSVAGSNDIPLDAIKHSRPGPTANGASKERTKAF-PNVK 1192
1011 PSEPADR--PCAGRCRPECNVKRFESAEDLESRYQVHRDLTPTSGPRKGAQTSVPKPK 1069
1193 NESPLNVKTEKVRDSTQ-----DELSKCVKRCG:VFLDEVVYALHM-----SCHDS 1240
1069 TYQCIKCOMTFENERE:QIHVANKMISEG:INHEKLCNQMFSPAKLLCHLIEHFEFGV 1128
1241 GPFQCSIC-----QHL-----CTKCYCFTHLQGLHRRNAQ 1272
1129 GTFKCPVCTVFVQANKLQCHIFAVHGGDEKTYQCSQCPKXFFFCLEQNH:MSQHAQ 1186

RESULT 5
Q9NZ13 PRELIMINARY: PRT: 1224 AA.
AC Q9NZ13: 094860;
DT 01-OCT-2000 (Trembl) 15, Created;
DT 01-OCT-2000 (Trembl) 15, Last sequence update;
DT 01-MAR-2003 (Trembl) 23, Last annotation update;
DE Smad and Gtf-interacting zinc finger protein (Fragment);
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN 111;
RP SEQUENCE FROM N.A.
RX MEDLINE20123442; PubMed10660046;
RA Hata A., Seoane J., Lagna G., Montalvo E., Hemmati-Brivanlou A.,
RA Massague J.;
RT Smad uses distinct DNA- and protein-binding zinc fingers in separate
RL Smad and Gtf signaling pathways.;
RL Cell 100:1229-243(2000);
DR ENBL; AF221712; AAF28354.1; ;
DR HSSP; P08C47; 1SP1.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF000096; zf-C2H2; 29.
DR ProDom; PD000023; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 27.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 21.
KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
FT CON TER 1224 1224
SQ SEQUENCE 1224 AA; 137949 MW; A2DE880C1AA925FC CRC64;

Query Match 3.9%; Score 267.5; DB 4; Length 1224;
Best Local Similarity 19.5%; Pred. No. 4.3e-09;
Matches 286; Conservative 150; Mismatches 537; Indels 491; Gaps 71;

QY 48 QSDAAELNKEEELH-----VQDP-----SSSKKQDKS--AVLSE--KAGNYESP 91
DB 13 QQFESLAULTDRAHRCPCGDDDPQLSNWASSPSSKDVASPTQYIGDGDLSJGEEEG 72
QY 92 SKGNFPSPHDEVTRNKLAFSFFAAGVCEPLKSPORAEADPPQWACTTPSGSJETK 151
DB 73 GTGLPYPC--QFCXKSTIRLSY-----LKRHEQHS-DKLPPKCTYCSRLFKHK 118
QY 152 ECKMSKPAETE:GCAQSGCANCQGLSPVSVASKAPQVPSGCVRLNKSXTDLLVNDPN 211
DB 119 RSPDRHKL--HTGCKKHYHCEEAAPS-----RSDHLKHLKTHSSSK 160
QY 212 PAFSLPELQCFKCNICGVGVGVCNDPTDL:KHF--KYHLGLNRT----- 254
DB 161 P-----FKTVCNRP--SSTSSLSHMQAHKXKHEHLAKSEKAKKRCDFMCDYCE 209
```

QY 255 ---ROAELDSKILALHNMVQFSHKQKQVN----- 283
Db 210 DTFSQTELEKHLVTRH--POLSEKADLQC7HCPEVFDENTLLAHIIHOAHANQKHKCPX 267
QY 284 ---RSVFSGLVODINSR-----PVL-----LNGYD--VQVTSST 315
Db 268 CPEQSSVEGVYCHLDSHRQDSSNHSVSPDPVLGVSMSMSATPDSSASVERGSTPDST 327
QY 316 ---FIGIGRKTDCOGNTK-YPRCKFCNFTYMGNSSTELQHFLOTHPNK1KA----- 364
Db 328 LKPLRGQKMRDDGGGKGTAVYVSCYCKRDF-NSLAVIEHLKT7HADKQCQGH7COIC 386
QY 365 ---SLPS-----SEVAKPEKSNKSNKIPALQSSD-----SGDLGKQCDXITVX-A 405
Db 387 L3SMPTLYNLNHRVRLKH--NHAYVPMQFNGNISAFHCNYPCEMPADINSLOEIRVSHC 444
QY 406 GDDTFVGVSVPIKPLDSSRQNGTEATSYWCKFCSCFSCSSSLKLLHYKQKQCAVOSG 465
Db 445 GPNA-----NPSDGNNAFFCNCQSMGF7TESS--L7EHIQQAHCVSQSA 486
QY 466 GLN-----PELNDKLSRGSVT-----NQNDL-----AKSSEGETMT 496
Db 487 KLESPVQVQTSFMEVSCYCTNSPIFGSILKLTKHKENHKNIPLAHSKSKAEQSPV 546
QY 497 KTDKSSGAKKXDFSSKGAEDNMVTS--YNQFCDFRYSKSHGPDVIVVGLJRHYYOOLH 554
Db 547 SSDVEVSPKQRLS---ASANSISNGEYPCNQCDLKFSNFES-----FQ7HLKLH 594
QY 555 NIHKCTIKHCPRCPLGCSPEKHLGEITVPPACKSN--CSHC-----ALLL 600
Db 595 LELLURKQACPOCKEEDFQESLLOH7VHYMTTSHYVCSCKQFSVDOLQKHLLDM 654
QY 601 HLPGAAGSRVHVKHCHQCSFTTPDQVLLPHYESVHESQAS-----DVQEAHNL 651
Db 655 H-----TFVLXHTLCOEVDFSKVSIQVHLAVKSNKXKVRCTACNWDPRKEA--- 703
QY 652 QGSDQQQSVKES-----KEHSCTKCDFTQVVEEISRYRAHSCYKRCOSFTAAQTQ 705
Db 704 ---DLQVHVXGSHLGNPAKAHKCIFCGETFTSEVELQCHITTHSKYKCKFCSKAFHAI 760
QY 706 SILLEPNTVHCOEQDIT7TANGEDG-HAIST7KEPKIDFRVYNLLTPDSXNGEPVSEV 764
Db 761 LLEKHLREKHC-VFAATENG7ANGVPPMATKAEP-ADLOQLLNKPEA---PNSH-- 812
QY 765 VKREKLEKDG7KEKVM7TESDDLL-RNVTWRGADILRGSPYTCAS7CLLTPVSGTQEQ 823
Db 813 -----EASEDDVASEPMYGCJIC--GAATVYEV7-----LQ 842
QY 824 7KTLRDSNPVAAHLARP7YGLAVETKGFLOGAPAGGKSGALPOQYPASGENSKSDESQ 883
Db 843 NHRLRD-----HNIRP-----GEDDGRKKAETIKSHKNCVSR 877
QY 884 SILRRRGSGVFCANCLITKTSLMRKNANGVYVQACGLYQKHLSTPRPLN7IKONGEQ 943
Db 878 TP-----PSENGLRHL7THRGPAK-HYMCPICG-----EFP7SLL7L7E7----- 917
QY 944 IIRRRTRKPLNPEALQAEGLNQKQSGSDEQVNGSFLERRSEDH7L7ESG7REIP7PSLSK 1023
Db 918 -----KVTHSKSLDTG7CRICKMP7C--SEEFIEHCQMH---PCL-- 953
QY 1004 YEAQSGLT--KHSQAQPV7VSGTLDIHK--RQOP7H7QIKSPQSB7GDPNGNSSVSSEK 1059
Db 954 ---RNSLTGFRVCVCMQ7V7STLELKIHTG7FMQKL-----AGSSAAS--- 993
QY 1060 GSSRGSP7E7KYRPAKH7N7SPGSP7E7KYQVPLG7L7PVHND7FQSEAD7K7L7R7F7SK7L 1119
Db 994 -----SPNGOGLQK-----LYKCA7L7CK7ER7SKOD7LV-----K7 1022
QY 1120 SVPGNPHYL-----SHV7PGL--PN7PCQNV7PY7PTFN7L7P7H7SAV7GSDND7I7PL7A 1167
Db 1023 DVNGLPYGLCAGMARSANGV7GLAPPEPADR--PCAGLR7PEC7SV7FES7ADE7LES7YQ 1080

QY 1168 IKHSPPGPTANGASKKTKAP-PNVKNBPLNVK7TEKVORSTQ-----DELSTYC 1217
Db 1281 V0HRDLTPT7SGPRKQTQ7SPVPRKX7YQ7CIKQMT7FENERET7QIHVANKY7FEGINHEC 1240
QY 1218 VHC7IY7PL7DEW7VAL7H7---SCHGDSG7PQCS7C-----QEL----- 1251
Db 1141 KLCNQMT7SAK7L7CHL7IEH7FEG7G7T7K7CPV7CTV7FVQANK7LQOH77FAV7HQCED7KI7YD 1200
QY 1252 ---CTK7YDFT7H7IQR7G7JHR7NNAO 1272
Db 1201 CSC7POK7FFOTELQ7N7FMSQ7HAO 1224
RESULT 6
QD Q8C1Q2 PRELIMINARY; PRT: 1311 AA.
AC Q8C1Q2:
D7 01-MAR-2003 (Tiemblure: 23, Created)
D7 01-MAR-2003 (Tiemblure: 23, Last sequence update)
D7 01-MAR-2003 (Tiemblure: 23, Last annotation update)
DE Ecotropic viral integration site 3.
GN EVI3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Warming S.; Jiu P.; Suzuki T.; Akagi K.; Lindtner S.; Pavlakis G.N.;
RA Jenkins N.A.; Copeland N.G.;
RT "Evi3, a common retroviral integration site in murine B-cell lymphoma,
RT encodes an EBFAZ-related kruppel-like zinc finger protein.";
RL Blood 3:0-3(2002).
DE EMBL: AY147426; AAN39819.1; -
SQ SEQUENCE 1311 AA; 147623 MW; 53810EC6345C3D37 CRC64;

Query Match: 3.91; Score 265.5; DB 11; Length 1311;
Best Local Similarity 19.24; Pred. NC. 6.5e-09;
Matches 302; Conservative 177; Mismatches 516; Indels 575; Gaps 82;

QY : MVKVKQPP7RNVA7EGEGQ7LEPIG7ESKVS7GK7KFS7AQQ7SEN7DOS7DAA7-----EJ 54
Db : M5SRKQAKP7SL-K7F7PK7LED7IED7GE7AV7CCK7PE7DGE7ELE7EED7AVH7S7C7LQ7VES7L 59
QY 55 NHKEER7SLH-----YQD-----PSSSK7DL7K7SAV7LSE7K7AG7F7N7Y7SP7SK7GN7P 98
Db 60 SD7TEH7K7L7HCCQ77CQ7GV7DED7P7SC7NP7ASS7P7SS7K7DT7SP7SH7GEG7CC7FEE 114
QY 99 S7PHD7-EV7DRN7LAF7P7PAAG7VCE7PL7KSP7RAE7AD7PQ7DMA7CT7P7SG7S7LE7TK7EQ7MS 157
Db 115 G7P7VQ7CF7CK7S7F7SL7SY-----LKH7EQ7SH7S7DK7L7P7FK7CT7Y7CS7L7FK7H7KR7SR7H 164
QY 158 P7ATE7T7G7QAQ7SC7AN7CQ7-----LSP7SV7ASK7NP-----QV7P7SG7V7RL7N7K7SD7LL 205
Db 165 IK7--H7G7DK7KY7HC7S7DAA7F7SS7S7HL7X7HL7K7HT7SN7K7PK7CA7V7CR7G7FL7SS7L7H7R7MQ 222
QY 206 VND7NP7DPA7P7L7SP7ELO7DF-----KCN7IG7Y7G7ND7PT7DL7IK7H7FY7H----- 247
Db 223 V7ERN7DGS7G7SG7F7ED7C7WK7M7K7CT7CK7S7C7Q7EE7GF--DF7ED7L7Q7H7AE7H7E7C7P7NE7D7RAA 280
QY 249 ---L7GL7NR7TRQ7DAE77DS7K7IL7AL7HN7VQ-----F7SH7SK7Q7K7Q7N7R7SV7P 287
Db 281 LQ7MY7CHE7L7F7VE7TS7LM7NH7IE7V7G7G7EK7KN7S7C7S7E7SL7T7VE7EL7Y7SH7MD7SH7Q7Q7E7SC7-- 318
QY 289 SG7V7QD7INS7RP7V7L7N7GT7Y7CV7TS7GG7TF7IG7I-----GRX-----TPD 325
Db 339 -----NH7SN7SP7SL7V7G7V7TS7SS7TP7D7SN7LS7V7DS7TW7E7AA7P7I7PK7S7GR7K7RA7A7Q7SD 392
QY 326 CQ7N7TK-----Y7PR7CK7F7CN7FT7MG7NS7STE7EQ7H7FLQ7H7---P7X7K7AS7L7PS7SE7VA7K7P7SEK 377
Db 393 MT7G7PS7SK7QA7KV7TS7C7I7Y7CN7KQ7L7F7SS7LAV7LQ77HL7K7TN7H7L7DK7P7Q7AH7I7CQ7Y7C7E7VL7PS7Y 450


```

Db      545  SPFEVSCPTNSPIFGSILKTKHKNKIP:AHKKSKAEBSPSSVSSPR 604
QY      508  KDFSSGAEDNMVTS--YNCQFCDFPRYSGHGDVIVWGFLRHYYQQLHN:HKCTIKHCP 565
Db      605  QRLSGSA--NSISNGEYPCNQCDLAFSNFES-----FQTHLKLHLELLRKQACP 652
QY      566  FCPRGILCSPEKHGEITYPACRKN--CSHC-----ALLLHLSPGAGSSR 611
Db      653  QCKEDFDSQESLLOHLTHVMTTSTHYVCESDKQFSSVDDLOKHLLDMH-----TF 704
QY      612  VRHOCHOCSTTPDQVLLPHFYESVHESQAS-----DYKQEAHLQSGDQCSVKE 662
Db      705  VLYHCTLCQEVDFSKUSIQVHLAVKSNKMYRCTACWDFRKA-----DLQVHVKH 758
QY      663  S-----KEHSTCKDPIITQVEBEISRHVRAHSCYKQCQSFQTAADTOSLHEHNTVHC 716
Db      759  SHLGNPAKAHKCIFCGETESTEVELOCHITTHSKKYNCRFCSPKAPHAVILLEKHLREKHC 818
QY      717  QEQDITANGEEDEGHAISTIKPEPKIDFRVYNLLTPDSKMGEPVSSEVVKREKLEKGL 776
Db      819  -VFDAANGTANGVPFTSTTKAEPADLOGLMLKNPEA-----PNSH-----859
QY      777  KEKVMTSESSDDL--RNVTWRGADILRGSPSYTOASIGLLTPVSGTQEQTKTLRDSNVBA 835
Db      860  -----EASEDDVASEPMYGCDC--GAAYTMEV-----LQNR-LD-----895
QY      836  AHLARIYGLAVETKQFLOGAPAGGKSGALPOQYPASGENKSKDEQSLLRRRGSGVF 895
Db      896  -HNIRP-----GEDGSRKAEFIKSHKNCVCSRTF-----F 927
QY      896  CANCLITKTSLRKKNANGGVNACGLYOKLHSTPRPLNI:KQNGEQIIRTRKRLNP 955
Db      928  SENGREHLQTHRPAK-HYMCPICG-----ERFPLLTLTBH-----964
QY      956  EALQAEQLKQKRGSENEQVNGSPLRRSEDLHTEHQREIPLPSLSKYEAQGSJ--KS 1013
Db      965  -----KVTHSKSLDTGCRICKNPLO--SEEFIEHCQM--PDJ--RNSLGFRC 1009
QY      1014  HSAQQPVLVSOTLIDHK--RMQPLHIQIKSPQESTGDPNGSSSVSEGGKSGRSPIEKY 1072
Db      1010  VVCMQTVTSLKLTGHGTTHNQKL-----AGSSAAS-----1040
QY      1072  MRPAKHPNYPSPGSPTEKYQPLFGLPFVHNDQFSEADMLRFSKYKLVGPHYL--- 1226
Db      1042  -----SPNGQGLQK---LYKALCLKEFRSKQDLVR-----JDVNGLPYGLCAG 1061
QY      1129  -----SHVPLG--PNPCQNYVPYPTFNLPHFSAVGSNDIPIDLAIKSRPGPTANG 1179
Db      1082  CMARSANGQVGLAPPEPADR--PCAGLRCPCECNVVFESAEDLESHQYDHRCLTPE-SG 1239
QY      1180  ASKEKTKAP-PNVKNGPLNVVTEKVDNSTQ-----BELSTKCVHCG:VFLDEVM 1229
Db      1140  PRGAQTSVPVRKTYQCIKCRMTFENEREIQIHVANHMIEGINHECKLNCMPDSPAK 1199
QY      1230  YALHM--SCHGSGSPQCSIC-----QHL-----CTDXVDEFT 1260
Db      1200  LLCHLIEHFEGMGITKPCVCTFVQANKUQCHIFAVHGQEDKTYDCSQCPQFFOT 1259
QY      1261  HIQRLGLHRNNAQ 1272
Db      1260  ELQNHTMSQHAQ 1271

```

RESULT 8

```

Q8VDS6
ID      Q8VDS6      PRELIMINARY;      PRT; 1311 AA.
AC      Q8VDS6;
DT      01-MAR-2002 (TrEMBLrel 20, Created)
DT      01-MAR-2003 (TrEMBLrel 20, Last sequence update)
DE      Hypothetical 147.7 kDa protein.
OS      Mus musculus (Mouse).

```

```

OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rosalia; Sciurognathi; Muridae; Mus;
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strausberg R.;
RT      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EXBL; BC021376; A422376.1; -.
DR      InterPro; IPR00345; CysC_heme_bind.
DR      InterPro; IPR07087; Znf_C2H2.
DR      Pfam; PF00096; zf-C2H2; 28.
DR      SMART; SM00355; Znf_C2H2; 30.
DR      PROSITE; PS00190; CTOCHROYE_C; 1.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 27.
DR      PROSITE; PS00157; ZINC_FINGER_C2H2_2; 22.
KW      Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ      SEQUENCE 1311 AA; 147665 MW; C8D922B5335C099A CRC64;

Query Match      3.9%; Score 264.5; DB 11; Length 1311;
Best Local Similarity 19.2%; Pred. No. 7.6e-09;
Matches 302; Conservative 177; Mismatches 516; Indels 575; Gaps 82;

QY      1  MVRKKNPPLRVVASEGEGQILEPIGTESKYSGNKKEFSADQMSQNTDQS-DAA-----EL 54
Db      1  MSRRKQAKPRSL-KDPNCKULEDKIEDSEAVCKKRPEDGGELEBEDAVHSCDSCLOVFESL 59
QY      55  NHKEEHSJH-----VQD-----PSSSKKDLKSAVLSKAGFNYESPSKGNFNP 98
Db      60  SDITEHIIHQQLTGDVDEDDPCSCSPASSPSKQDQTSFSGEGCDPGEB---EGG-P 114
QY      99  SFPHD-SVTDNM-AFSPPAAGGVEPLKSPQRAEADDPQDMACTPSGDSLETKEQKYS 157
Db      115  GLPYPCFCDFKSFRLSY-----LKHGEGHS-DKLPFKCTYCSRIFFKHRSRDRH 164
QY      158  PKATEETGQAQSGAGANQCG-----LSPVSVASKNP---QVPSGGVRLNKEKTDLL 205
Db      165  IKL-HGCKKYH-SECDAAFSRSCHLKIHKHTSNKPYKCAVCRGFLSSSSJHGMQ 222
QY      206  VND-NPDPAP-SPPELQDF-----KCNIGYGYGNDPTDLIKHFRKYH-----247
Db      223  VHEANKDGSOGSRMEDKMKMDTKCSQCEGF--DFPELQKHAECHPSCSPNEEDRAA 280
QY      249  ---LGLNRTQDAELFESKILALNMVQ-----FSHKDPQKVNRSVP 287
Db      281  LQMYCHELFVEETS-LNHHIEQVHGGEKKNKSCS-CSESF-TVEELYSHMDSHQPESC-- 338
QY      298  SGV-LQDINSRPVJLN-GTYDVQVTSQGTFIGI-----GRK-----TPD 325
Db      339  -----NHSKSPSVTVGYTSVSSTTFCSNLVSDSSTMVZAAPIPKSRGRKRAAQOTSD 392
QY      326  CQGNK-----YPRCKPCNFTYKGNSTTECHPLQTH---PNKIKASLSPSEVAKPSEK 377
Db      393  MTGPSSKQAKVTYSCLYCN-KQFSSLAVALQHLKTMHLCKPEQAH-CQVCLEVL-PSLY 450
QY      376  MSNKSIAL-QSSSG-----DLQKQCKITVRAQDDTP-----41C
Db      451  KLEHLKQVHEAQPGG-VSAWPAIVYQCNFCSEVNDLNTLQEHTRCSRSGFANPAKDS 510
QY      411  -----VGYSVPKPLDSSRON-----GT---EATSYWCKFC 439
Db      511  NAFFCPHCYMGFLT-----DSSLJEHTRQVHCD-SGSRFGSPVLGTKEPWEVYSCSYC 565
QY      440  SPSCESSSLKLEHYQKHQGVQSGGLNPELNDKLSRG-----SVINQDL-----486
Db      566  TNSPIFNSVLKLNKH:KENHKNPL-ALNYIHNGKKSRLSPLSPVA-EQTTL-KQMGTVG 624
QY      497  ---AKSSEGETMTKTDKSSSGCAKKDFSS-----KGAEDNMVTSYNCQCFDPRYSKSHGP 538
Db      625  GGPARRAS-GEYIC-----KQCGARYSLDSSFQTHLKHLDLTVLPKLTCPQCKKEFPNQS- 678
QY      539  DVI-VVGPPLRHYYQQLHNHKT-KHCPFCPRGLCSP---EKHLGEI-TYFPAKRSKNSH 594
Db      679  -----LLXKV-TIHEKITSTYVICESCCKQFTSVDDLOKHLLDMHTFVF-----FRCTL 726

```

```

595 QY C-----ALLLHLSPGAAGSSRVKHQCHQCQSF--TPDQDVLLPHVSVHESQASVYKQE 647
727 DB CQVFDSKYSIUHLAVHSHNEKKV-YRCTSCNMDFRNE--DLQJH-----VKH- 773
648 QY ANHLQDGSQGVKESKEHSTKC--DFITQVEE--ISRHYRAHSCYKRCOSFTAA 703
774 DB -NHEE-----NQGVKHKICFGSGFTGVELQCHI--THSK-----YNCRFCSKAFHA 820
704 QY TQSLLEPHFNVHQEQDIT--TANGEBDGHAI--KEEPKIDFRVYINLTPDSKVGEPVSES 763
821 DB VILLEKHLREKHC-VFETKTPNCOTNCASQVQKEAELO-----TLLEN----- 864
764 QY VVKBREKLEEDGLKEKXVWTESSDDLNVNWRGADILRGSPSTQASLGLLTFSVGTQEC 823
865 DB --SQESHNSHDGSEEV--DSSEPM-----YCCDIC--GAAYTMETL-----IQ 902
824 QY TKTJRDSPNVEAAHLARPIYGLAVETKGFQGPAGGSEKSGALPQQVPASGENKSKDESQ 883
903 DB NHQLRD-----HNIRP-----GSAIVAKKA 923
884 QY SLRRRRGSGVFCAACLTTKTSLRKNVANGYVNCVACGLYQK--HSTPRPLNI--KQNGEQ 943
924 DB EJHK-----GNYKCNVCS-----RTFSENG-- 944
944 QY IIRRRTRKRLNP-----EALQAEQJNKOQRGNEECVNGSPLERRSED 986
945 DB -LREHWOTLGPVKHYKPCIGERFPFSLTLTEHKVTHSKSLQTKGRICKXPLO--SEE :001
987 QY HLTEHQREIPLPSLSKYEAQGSJTKSHSAQOPVLVSGTLD--HKRMOPHLIQ--KSPQEST :046
1002 DB EFLEHCOMH--PDL-----RNSLTGFRVCVCMQTVTSTLEL-KHGTTFHQ-----:044
1047 QY GDPGNSSSVSEKSGSSRGSPIEKYMRPAKHNPSPGSPIEKYQPLFGLPFVHNDFQS :106
1045 DB -KTGNGSSVQ---TTGRGQHVK-----LYKASCLEKFRS :076
1107 QY EADWLRFWSKYKLSVPONPHYLSHVPGLPNPCQNY--VPYPTNLPHPSAVG-----:1157
1077 DB KQDLV-----KLDINGLPY-----GLCAGCVNLSKSSGPGSLPPGASRPGLGQNESL :124
1158 QY -----SDNDIP--DLAIKHSRPGPTANGASKETKAPNVKN :1193
1125 DB SAMEGKGKAGLKTRCCSNVKESESELQNHQTVHRELVPDANKTQKTPQVSP-----:1180
1194 QY EGPLNVVYKTEKVRSTODELST-KCVHCGIVFLDE-----VYVALHN-----:1234
1181 DB --MPRSPSQSDEKTYOCIKCMVFYNWDIQVHANHM--DEGLNHECKLCSQ :1232
1235 QY -----SCH-----GSGPQCSIC-----OHL-----CT :1253
1233 DB TFDSPAKQOCHLHESFEGMGG--FKCPVCFTTVQVANKLOQHIFSAHQGECKIYDCTQCP :1292
1254 QY DKYDFTTHIQ 1263
1293 DB QKFFPTELQ 1302

RESULT 9
Q925J5
ID Q925J5 PRELIMINARY; PRT: 1669 AA.
AC Q925J5;
DT 01-DEC-2001 (TReMBUrel. 19, Created)
DT 01-DEC-2001 (TReMBUrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBUrel. 23, Last annotation update)
DE Zinc finger 142.
GN ZFP142 OR ZNF142.
OS Mus musculus (Mouse).
OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCB_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]

```

Db 773 SFDS-IETPALVAERE-----:|||:PVVEKLASEPPRN 601
Qy 934 NIKQNGEQIIRRRTRKRLN-DEALQAEOLNQKQGSNEEQVNGSFLERRSE-----985
Db 802 PLISEEAPNTFFKAALTAETVPLPPFSESELKAYRRQDKECAELVLSGRVQMVIQGE 862
Qy 986 -----DHLTESHQREIPLPSLSKYRAQSLTKSHSAQOPVLVSQTLDIHKMQPL--HI 1037
Db 862 GRAFRCPHCFPIIRREKALTTHSKSCQG-----RREPLLCPEGCAGFKOORGLSTHM 914
Qy 1038 QIKSPQESTGCPGNSSSVSGKSSRSGSPIEKYMRPAKHP-----NYSPPGS 1085
Db 915 MKKCPV-----LLKNKALPKPVSPTLHPQLPDNQASQDAESKRPPPL 957
Qy 1086 PIEKYQVPLFGJPFVHNDFOSEADWLFWSKYKLSVPGN-PHYLSHVPGJPNPCQNYVPY 1144
Db 959 P-----SKVELLPKDAFSDLPGGPGVVEEPLPTFSDF 989
Qy 1145 PTNLPHPSAVGSDNDIPDLDAIKHSRPGPTANGASKETKAPPNVQNEGLNVVVKTEK 1204
Db 990 PT-SPP-----ENSLP-----TGTSK-----1005
Qy 1205 VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHGSGPFCSCIOHLCTDKYDFTHIOR 1264
Db 1006 -----FHFOGKFKHCSSCTFLCSRLSLISHSVTE 1034
Qy 1265 GLHRNNAQVKNKPK 1280
Db 1035 GCRGGGQKRGKRPQ 1050

RESULT 10
Q8UVK1 ID Q8UVK1 PRELIMINARY; PRT: 589 AA.
AC Q8UVK1;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Ikaros-like transcription factor IKLF1.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hypercarlia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OC NCBI_TaxID=7757;
RN [1]
RE SEQUENCE FROM N.A.
RA Mayer W.E., O'Ruigin C., Terric J., Saraga-Babic M., Tichy H.;
RT "Identification of two Ikaros-like transcription factors in lamprey.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF424733; AAL67302.1;
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Znf_C2H2_5.
DR SMART: SMC0355; Znf_C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 4.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 589 AA; 63886 MW; 90C00A5C3167F61A CRC64;

Query Match
Best Local Similarity 19.4%; Pred. No. 3.2e-09;
Matches 143; Conservative 62; Mismatches 190; Indels 343; Gaps 29;

Qy 603 SPG-----AAGSRVHKHCQCFTTDPDVLFLPHVESVHESQASDVKO-EANKHLOGSDG 656
Db 117 SPGETRFSSGSSKLAADVCGMVICGNVLVWVKRSHTGERPPQSCGASFTQXGL 176
Qy 657 QQSVK-----ESKHSCTKCFITQVEEISRYRAHSC---YKROCSTFAADTQSLLEH 710
Db 177 VRHKLHTDEKPFKCHLCSACRRDALMGLH-RTHSVGPKYKCHSCRCYKORSLEH 235
Qy 711 FNVHCOEQDITTANGEDGHAISTIKBPKIDFRVNLTPDSKVGPEVSVVKREK 770
Db 236 LERCPAYCOQLSTRNQEDA-----DIRVH-----MDQVPA-----266

```

```

Qy 77: EKDCLKEKVMTESSDOLRNVTWRGADILRGSPSYTOASLGILLTPVSGTQCTKTLRDS 830
Db 267 ---DGLLE-----271
Qy 831 PNVEAAHLARPIYCLAVETKGFQCAPAGCKEKGALPCQYPASGENKSKDESOLLRRR 890
Db 272 ---AGSDRI-PLPNQLPG-----SFLAKRK 292
Qy 891 GSGVFCANCLTTKTSARKNANGVCNAGLYQKJHSTPRPLNIKONNGEQIIRRRTR 950
Db 293 S-----STPKL-----FNQ 302
Qy 951 KRJLPEALQAEALNKQKQSGNEBQVNSPLERSEDLTSHQREIPLPSLSKYKQSL 1010
Db 303 KRLN---NCLSDIRCEQWSSDO--NEAP-----HFVEQ-----AMSTYLGVRPL 342
Qy 1011 TKSHSAQOPVLVSQTLDIHKMQPLH-QIKSPQESTGCPGNSSSVSGKSSRSGSPIEK 1070
Db 343 LSQSPSQPLVLSHSES--NRIOF-----PSFLAIP-----372
Qy 1071 YMRPAKHPNPSPPGSPPIEKYQPLFGJPFVHNDFOSEADWLFWSKYKLSVPGNPHYLSH 1130
Db 373 -----KPPNS-IEDYA-PVIGAVY-----SH 392
Qy 1131 VPGJPNPCQNYVPPTNLPH-PSAVG-----SDNDIPDLAIKHSRPGPTANGASKEX 1184
Db 392 TVGQSPSPRSQSP-----LPCHGFCSSGKQGTSEQLPRDRASLGFHGRAAAASPS 446
Qy 1185 TKAPPNVKE-----GPNVYVKEK---VDNSTCDELST-----1215
Db 447 NSCPSTETESSHEERORLFRANCSAQPTGNSSATQDAASTPARADVQLGEGG 506
Qy 1216 -----KCVKCGIVFLDEVMYALHMSCHGSGPFCSCIOQ 1249
Db 507 QPRPPASEGGAYRVMGEDGEYMRAYNCLHCQVIFLDHVMYTLHNGCHGRJDPFCNVCG 566
Qy 1250 HCTCKYDFTHIORGJH 1267
Db 567 HRSRDYEFSSHIRGEH 584

RESULT 1:
Q8ESD2 ID Q8ESD2 PRELIMINARY; PRT: 1267 AA.
AC Q8ESD2;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Early B-cell factor associated zinc finger transcription factor.
GN EBF4Z.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RA Croci L., Corradi A., Vauti F., Wurst W., Roschi X., Gonzalez G.G.;
RT "CNA sequence and map assignment of Ebf4z, orthologous to the zinc
RL finger transcription factor gene Roaz."
EMBL: AF188609; AAG17053.1;
DR HSP; P08347; 1SP1
DR MGC; MGC11891217; Ebf4z.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Znf_C2H2; 28.
DR PRODOM: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 26.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 27.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 21.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1267 AA; 131578 MW; 6EB8FA77-E0CDF3E CRC64;

```

Query Match 3.8%; Score 262; DB 11; Length 1167;

Best Local Similarity 19.1%; Pred. No. 9.5e-09;

Matches 261; Conservative 138; Mismatches 512; Indels 456; Gaps 60;

```
QY 27 ESKVSGKKEFADQYSENTDQSDAAELNHKEHS:HVQDPSSSSKKL-----75
DB 136 KSEKAKKDDFCDCYCEDTFSQTELE---KHVLTLLH---POLSSKADLQCHQPEVEVD 189
QY 76 KSAVLSE---KAGFNYEYSPSGKGFPPSP-----PHDEVTDKXLMAFSPFAAGGYCE 123
DB 190 ESTLLAHIIHQANQKXKCPMEQPEQSSVEGYCHLDHRQDPSNNHVSVPDPVGSVAS 249
QY 124 PLKSPORAEADPDQDMACTPDSGDSLETKEQKMSPRATEETGQAQSGQANCGG-SPYSVA 183
DB 250 -----MSSATPDSSASVERGSTPDTLKLPLRGKKQMRCDQSW-----287
QY 184 SKNPQVPSDGGVRLNKSRTDLLV-----NENPDPAFLSPQLDFKCNICGYYGYN 234
DB 288 ---PKVYVSCPYCSKRDFTSLAVLE:HLKTIHADKQP-----QSHTCQIC---LDS 332
QY 235 DPT--DLIKHFRKYHLGLHNRTRQDAELDSKILALNNMVQFSHSDKFQKVRNVSFGVLQ 292
DB 333 MPTLYNLNHRVKLHKS-----HAYPVMQFCNISAF---HCNYCPSEYFA 373
QY 293 DINSRPVLLNGYDVQVTSGGTFTIGIKRTDCCQNTKYFRCKFCNFTYMGNSTELEQ 352
DB 374 DINS-----LQEHIRVSHCPNANPPD--GNNAFF--CNQCSMGFLTESS--LTE 417
QY 353 HF-QTHPNKIKASLPSESEYAKSEKSNKSI:PALQSSDGLGKQKQIDTKVAGDUTPVG 412
DB 418 H:QOAH-----CSVGSKLESP-----434
QY 413 YSVPIKPLDSSQNGTEATSYNWKFCSPSCSSSLKLEHYGKHQAVGSGG:NPSELN 472
DB 435 ---VQPTQSFME-----VYSCPYCTNSPIFGSILKLT:KHKENHKKI-----PLAH 478
QY 473 DKLSRGSVINQDLAKSSGE:WTKTDKSSGAKKKDFSSKAEONMVT--YNCQCFDF 530
DB 479 SK-----KSKAEQSPVSDVEVSFKQRLSGSA---NSISNGSYPCNQCTL 522
QY 531 RYKSKSGPDVIVVGLRHVYQLHNIHKTIKCPFCPRGLCSPEKH:GEITYPACRKS 590
DB 523 KTSNFES-----FQTHKLHLELLRLKQACPOCKEDFSQESL:QHLTVHMT-ST 573
QY 591 N--CSHC-----ALLLHUSPGAAGSSRVKHQCHQCSFTTPDVVLLHFHVESV 636
DB 574 HYYCESCDXQFSVDQLKXHLDMF-----TFVLYHCTLCQVFSKVSIGVLEJAVK 625
QY 637 HESCAS-----DVKQEAHNLQSGDQSQSVKES-----KESCTKCDFTTCVEEBI 681
DB 626 HSEKKYMYRCTACNWDPRKEA-----DLQHVKHSHLGNPAKAHKCFCGETFTSTEL 679
QY 682 SRHYRRAHSCYKRCQCSFTTAADTQSLLEHFNTHVCOEDITFANGEEDGHAISTIKPEPK 741
DB 680 QCHITTHSKYKNCRCFSKAPHAVILLEKHLREKHC-VFADAENGANGVPPSTKKAAP 738
QY 742 IDRVVNLITPDCKMGEPIVSESVVREKLEKDKLEKVTWTESSDDL-RNVTWRGAIL- 800
DB 739 ADUQGLMLKNPEA---PNSH-----EASEDDVDJASEPMYGCIC 774
QY 801 RGSFSYVQASGLLTPVSGTQEQTKTLRSPNVEAAHAE:VGLAVETKGLQAPAGG 860
DB 775 --GAAYTMEVL-----LQNHRLD-----HNIRP-----G 797
QY 861 EKSGLAPQOYPASGENKSDQESQ:LRRRRGSGVFCANCLITKSLWRKNANGGYVNAC 920
DB 798 EDDGSRKKAZF:KGSKCNVCSRTF-----PSENGLEHRLQTHRGPAK-HVMCP-C 847
QY 921 GLYQKUHSTPRPLNI:KONNGEOI:IRRTKRLNPEALQAEQ:LNKOORSENEQNGSPL 980
DB 848 G-----ERFESLLTLEH-----KVTHSKSLDTGTCTICKMPL 860
QY 981 ERSSEDLHTEHQREIPLPS:SKYEAGSLT--KSHAQOPVLVSQTLDIHK--RYQPLH: 1036
```

RESULT 12

```
Q925J4 PRELIMINARY; PRT: 1669 AA.
ID AC Q925J4
DC 01-DEC-2001 (ITEMB:rel. 19, Created:
DT 01-DEC-2001 (ITEMB:rel. 19, Last sequence update:
DT 01-MAR-2003 (ITEMB:rel. 23, Last annotation update)
DE Zinc finger 142.
GN ZFP142 OR ZNF142.
CS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCB TaxID:10096;
RN 1.1.
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RA Bringer M.A., Thompson J., Conroy C., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela C.M.;
RT "High-Throughput Sequence Identification of Gene Coding Variants
RT within Alcohol-Related QTLs";
RJ Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EXBL; AF132092; AK56118.1;
DR MSD; MG11924514; Zfp142.
DR InterPro; IPR003045; Cytochrome_b1nd.
DR InterPro; IPR003045; Cytochrome_b1nd.
DR Pfam; PF00096; Zf-C2H2; 30.
DR SMART; SM00355; Zf-C2H2; 36.
DR PROSITE; PS0190; CYTOCHROME_C_1.
DR PROSITE; PS0290; ZF_MHC_1.
DR PROSITE; PS0026; ZINC_FINGER_C2H2_1; 17.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 19.
KW Metal-binding; Zinc; Zinc-finger.
SC SEQUENCE 1669 AA; 186763 MW; 7D5138424A5198F6 CRC64;
```

Query Match

Best Local Similarity 18.4%; Pred. No. 1.7e-08;

Matches 213; Conservative 124; Mismatches 360; Indels 459; Gaps 50;

```
QY 366 LPSEV-----AKPSEKSNKSI:PALQSSDGLGKQKQ:TVKAGDTPVGVSV 415
DB 113 LPSQELLPAKLPFGHREPSEAST-PLPQESABEED-ABEESVTKDSQKV-YDKSQ 169
QY 416 PIKPLDSSRQNGTEAT-SYWKCKFC-----SFSCEI-----SSS 448
DB 170 GACQLEGRVSGTGESLFKTHYCPCKRCFKK:HLVSH:HLHPPDPSLQCPCKRFFTSK 229
QY 449 LKLEHYGKHGAVCGSG:LNPELNDKLSRGSVINQNDJAKSSEGETMTYTKDSSSGAKKK 508
```

Db 230 SKLTH-----LLRELGEKAHRCPLCHYSAVERNALNRHMASHEHEDIS----- 272
Qy 509 DFSSKGAEDNMVTSYNCQFC--DFRYS-----KSHGPDVIV-----VGP 545
Db 273 NFYS-----DTYACPVCRBEFRLSQALKHEKLSHTAAABEPLPLHCFQEGGTYVAP 324
Qy 546 ----LLRHVQOHLNI-----HK-----CTIKH-- 563
Db 325 DRKAFUKLKEITHGVRAVECRHHSCPMJFATAEAMEAHKSHYAFCHPCDPCFACSKNLF 384
Qy 564 ----CPFRGLCSP--EKLGEI-TYPPACRKSNC-----SHCALLJL 600
Db 385 RKXKKGHGPGBELRTPCFATFNVAOYDVGKMHAYEKIHQCSECNFATAHKLVLIR 444
Qy 601 HSPGAAGSRVKGHCQCSFTTPVDVL-----LPHYESVH 637
Db 445 HM--LLHTGERPKCELCDFTCROVSYLSKMLTHSNTKDYMCCTCGVTKWJLVS 501
Qy 638 -ESQASDVQBEAN----HLOGSDGQSVK---ESKEHSTCKCDFITQVEEELSRHYRAH 689
Db 502 MRKHAGDLRYQCQCSYRCHRADQLSSHLRHQKSLMCEVCAFAKRYELQKHMASH 561
Qy 690 ----SCVKRCRCSFTAAQTOSLLEHN-----TVHCQODITTAN----- 725
Db 562 HGTGAPLPICRYCSYQSRHKQALLSHENCKTHLREFHCALCDYRTFSNTTLFPHKRV 621
Qy 726 ----GEE-----DGHASTIKEPKIDFRVYNLL---TPDSK 755
Db 622 HGVMPCDQWQFCNASQEL-EGARQCLAPSDSGPSSQLSAQPEREDREHEIVANSNMQA 681
Qy 756 MGPVSVESVVRKLEZEKLEKVKWTBESSDDLNVNTRGADILRGSPSYQASLGJLT 815
Db 682 LPTNEBAGPKQ-----DGIE-----APQEDQVD-----SPSLGEVEEG-- 717
Qy 816 PVSGTQEQTKLRD--SPNVEAAHLARPIYGLAVETKGLQAGAPGKSGALPQCYPAS 873
Db 718 --GCTILHLEALRVELEPETEPLBELTETATVEFPLDPGFLGTERPGGLE--EAL 772
Qy 974 GENKSDBSOSLLRRRGSGVFCANCLTTKTLWRKNANGVYVGNAGLYQKLHSTPRL 933
Db 773 SSFDS-IETPALVAEE-----PVVEKLASEPERN 801
Qy 934 NIKQNGEQIRRTKRLN-PEALCAEOLNKQORGSNEEQVNGSPLEERS----- 985
Db 802 PLISEEAPNTFKALTAETVPLPFPFSESLKAMRRQDKEAELVLEGRVQMVVIOGE 861
Qy 986 ----DHLTESHQREIPLPSKYBAQSLTKSHSAQQFVLVSQTLDIHKMQPL--H 1030
Db 862 GRAFRCPHCPFI--REKALT--HSGSCQG-----RREPLLCEGASFKQORGLSTHM 914
Qy 1038 QIKSPQESTGDPGNSSVSEKSGSSRGSPIEKYMPPAKHP-----NYSPPGS 1085
Db 915 MKKCPV-----LLKNKALPKVPSPGLHPQLPDNOASQDAESRKPPL 957
Qy 1086 PIEKYQVPLGFLPFVHNDFOSEADLRFMSKYKLSVPGN-PHYLSHVPGLPMPQCNVY 1144
Db 958 P-----SKVELLPKAPSDLPDGPQGVSEPLTPSDP 989
Qy 1145 PTNLPHPHSAVSGDNDIPDLAIKHSRPGPTANGASKETKAPPNVKNEGLNVYKTEK 1204
Db 990 PT--SPP-----ENSLP-----TGTSEK----- 1005
Qy 1205 VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHDGSPPQCSICQLCTDKYDFTTHIQ 1264
Db 1006 -----PHEQCKFHCSCTCLCSRLSITSHTVE 1034
Qy 1265 GLHRNNAQVEKNGKPK 1280
Db 1035 GCRGRCQKRGKRPQ 1050

ID O93581 PRELIMINARY; PRT: 537 AA.
AC O93581;
DT 01-NOV-1998 (T-EMB:rel. 08, Created)
DT 01-NOV-1998 (T-EMB:rel. 08, Last sequence update);
DT 01-MAR-2003 (T-EMB:rel. 23, Last annotation update);
DE Ikaros.
GN ZNFN1A1 OR IKAROS.
OS Brachydanio rerio (zebrafish); (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spine;
RA Anemiyu C., Kawasaki H.;
RT "Characterization of zebrafish ikaros, a gene necessary for
RL differentiation of the immune system.",
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF092175; AAC61763.1; -;
DR ZFIN; ZDB-GENE-990526-304; znf1a1.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR007087; znf_C2H2.
DR Pfam; PF00336; zf_C2H2_5;
DR ProDom; PD000003; Znf_C2H2_2;
DR SMART; SM00355; Znf_C2H2_6;
DR PROSITE; PS00179; AA-TRNA-LIGASE 1; 1;
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5;
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 4;
KW Metal-binding; Zinc; Zinc-finger;
SQ SEQUENCE 537 AA; 58665 MW; 38F047938CF39A37 CRC64;

Query Match: 3.8%; Score 260; DB 13; Length 537;
Best Local Similarity 21.0%; Pred. No. 4,1e-09;
Matches 118; Conservative 71; Mismatches 202; Indels 172; Gaps 21;

Qy 914 LTVSGTQEQTKLRSP---NVEAAHLARPIYGLAVETKGLC-----GAP 857
Db 38 LSASTGJOHNRT--DKPLACNIKVEARDEENGLSCVMXGEAECAEDLILDSGAK 95
Qy 858 AGREKSG--ALPQCYPASG-----ENKSKDSQSLLRR 888
Db 96 VNASHAFDCKPAAYPAGAIRLPLNGKLCDDICGIVCIGPVNLMVTKRSHTEKSVLEQ 155
Qy 889 RRGSGVF---CANCUTTKSLWR---KVANGVYVNCAGL-----YQKJHS-P 930
Db 156 QKGERPCCQCCGASFTCKGNLLRHKLHSGEKPFKCHLCNVACRRRDALTGLRTHSVG 215
Qy 931 RELNI-----IKQNGEQIIRRTKRLNPEALCAEOLNKQORGSNEEQVNGSPLEERS 984
Db 216 KPHKCAVCGRSYKQRSLSBEKERCNYLQCMGLQNSIYTVKEENSONQREMPASERA 275
Qy 985 -----EDLTESHQREIPLPSKYBAQSLTKSHSAQQFVLVSQTLDIHKR 1031
Db 276 LVLDRIANNVAKKSNPQRFVGNELSELSPESGSGELMQPHVIDCAINSATSYLGAES 335
Qy 1032 MQPLHIQKSPQESTGDPG-----NSSSVSEKSGSSRGSPTEK--YKPPAKHPY 1080
Db 336 LRPL-----VQSPSACMVVSVFLYNLHKSTQAEKGVSAKDSAAAEHLILSKSKSASV 389
Qy 1081 SPQSGPTEKYQVPLGFLPFVHNDFOSEADLRFMSKYKLSVPGNPHYLSHVPLGNPCQN 1140
Db 390 DKQSGP-----SPSQDSDTDIESN-----NEERSAGVSGT----- 419
Qy 1141 YVPYFTNLPHPHSAVSGSNCLPLCLAAKHSRPGPTANGASKETKAPPNVKNEG----- 1195
Db 420 -----AAATGG-----LIVLTNHNMAPGMNGGL-----PQVKEEQCRHPE 453
Qy 1196 -----PLNVVKTE--KYDRSTQDEL-STKCVHCGIVFLDEVMYALHMSCHDGSPPQC 1245
Db 454 ALRAAGMDLSIASSEGEKFLSGDGEELRAYRCHICRVLFCHVWYTHIMGCHGFRDPFEC 513
Qy 1246 SICGHLCTDKYDFTTHIQRLHR 1268

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 29, 2003, 11:44:12 ; Search time 236 Seconds
(without alignments)
8418.074 Million cell updates/sec

Title: US-09-702-216-1_COPY_500_5000
Perfect score: 4501
Sequence: 1 atgtttctgtgtctctga.....agaaaaatggcgccctctca 4501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/2/ina/6CTUS_COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	1.4	1456	3	US-09-037-135-1
2	58.6	1.3	1984	4	US-08-733-622C-1
3	57.4	1.3	1004	1	US-08-465-590-8
4	57.4	1.3	1004	3	US-08-283-300A-7
5	57.4	1.3	1004	3	US-08-711-417C-8
6	57.4	1.3	1004	4	US-08-733-622C-21
7	57.4	1.3	1004	5	PCT-US95-09345-7
8	57.4	1.3	1386	1	US-08-465-590-3
9	57.4	1.3	1386	3	US-08-283-300A-2
10	57.4	1.3	1386	3	US-08-711-417C-3
11	57.4	1.3	1386	4	US-08-733-622C-16
12	57.4	1.3	1386	5	PCT-US95-09345-2
13	57.4	1.3	1551	3	US-08-711-417C-165
14	57.4	1.3	1611	5	PCT-US93-08743-3
15	57.4	1.3	1788	4	US-09-435-327A-19
16	57.4	1.3	3629	4	US-09-435-327A-18
17	53.8	1.2	2897	2	US-08-927-394-1
18	53.8	1.2	2897	4	US-09-016-434-1163
19	53.2	1.2	7218	1	US-08-232-463-14
20	50.6	1.1	1128	1	US-08-465-590-7
21	50.6	1.1	1128	3	US-08-283-300A-6
22	50.6	1.1	1128	3	US-08-711-417C-7
23	50.6	1.1	1128	4	US-08-733-622C-20
24	50.6	1.1	1128	5	PCT-US95-09345-6
25	50.6	1.1	1170	1	US-08-465-590-6
26	50.6	1.1	1170	3	US-08-283-300A-5
27	50.6	1.1	1170	3	US-08-711-417C-6

28	50.6	1.1	1173	4	US-08-733-622C-19
29	50.6	1.1	1173	5	PCT-US95-09345-5
30	50.6	1.1	1296	1	US-08-465-590-4
31	50.6	1.1	1296	3	US-08-283-300A-3
32	50.6	1.1	1296	3	US-08-711-417C-4
33	50.6	1.1	1296	4	US-08-733-622C-17
34	50.6	1.1	1296	5	PCT-US95-09345-3
35	50.6	1.1	1788	1	US-08-465-590-2
36	50.6	1.1	1788	3	US-08-283-300A-1
37	50.6	1.1	1788	3	US-08-711-417C-2
38	50.6	1.1	1788	4	US-08-733-622C-15
39	50.6	1.1	1788	5	PCT-US93-08743-2
40	50.6	1.1	1788	5	PCT-US95-09345-1
41	50.6	1.1	2049	1	US-08-465-590-5
42	50.6	1.1	2049	3	US-08-283-300A-4
43	50.6	1.1	2049	3	US-08-711-417C-5
44	50.6	1.1	2049	4	US-08-733-622C-18
45	50.6	1.1	2049	5	PCT-US95-09345-4

ALIGNMENTS

RESULT 1
US-09-037-135-1
; Sequence 1, Application US/09037135
; Patent No. 6,07034
; GENERAL INFORMATION:
; APPLICANT: Weigel, Ronald
; TITLE OF INVENTION: GATA-3 EXPRESSION IN HUMAN BREAST
; TITLE OF INVENTION: CARCINOMA
; FILE REFERENCE: SCR-76P
; CURRENT APPLICATION NUMBER: US/09/037,135
; EARLIER FILING DATE: 1998-03-09
; EARLIER APPLICATION NUMBER: 09/037,135
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1456
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(1451):

QV	2759	GGAGAAACAAAGTCGCAAGGATGAATCCAGTCCTCTTACGAGGCGTAGAGCTCCGGT	2819
DB	1005	GGACAGAACCGGCGCCCTCATTAAGCCCAAGCAAGGCTGTCTGCACCCAGGAGCAGGG	1564
QV	2819	GTTTCTTGCGCAATTCCTGCACCAAGACATCTCTCGGCGAAGATCAATGCG	2878
DB	1665	ACGTCTCTGCGCACTGTTCAGACCCACCAACCACTCTGGAGGAGGATGCCAATGG	1124
QV	2879	GGATATGTATGCAACGGGTGTGGCCCTTACAGAACTTCACCTCGACTCCAGGCTTTA	2938
DB	1125	GACCTCTCTGCAATGCTGTGGGCTCTACTACAAGCTTACAAATATTAAACAGACCCCTG	1184
QV	2939	AACATCATTTAAACAACAAC	2959
DB	1185	ACTATGAAGAAGGAGGCATC	1205

RESULT 2
US-08-733-622C-1
; Sequence 1, Application: US/08733622C
; Patent No. 6528634
; GENERAL INFORMATION:

```

Query Match      2.31; Score 57.4; DB i; Length 1004;
Best Local Similarity 60.64; Pred. No. 3 se-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

[illegible]

RESULT 4
US-08-283-300A-7
Sequence 7, Application US/08283300A
Patent No. 6172279
GENERAL INFORMATION:
APPLICANT: GEORGIOPOULOS, Katia A.
TITLE OF INVENTION: CAROS TRANSGENIC CELLS AND ANIMALS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, Suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Ascii text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/283,300A
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/238,212
FILING DATE: 02-MAY-94
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/121,438
FILING DATE: 14-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,695
REFERENCE/DOCKET NUMBER: MSP-827
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, Suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09345
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/283,300
FILING DATE: 29-JULY-94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,212
FILING DATE: 02-MAY-94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/121,438
FILING DATE: 14-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,695
REFERENCE/DOCKET NUMBER: MGP-027PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1004
PCT-US95-39345-7

Query Match 1.3%; Score 57.4; DB 5; Length 1004;
Best Local Similarity 60.6%; Pred. No. 3.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 2;
QY 3785 AATGTGTGCACCTGGCGATTGCTTTCTGGATGAAGTATGCTTTGTCATATGAGT 3844
DB 832 AATGGCAACATGCCGGGTGCTCTTCTGGATGACGTCATGTACACCATCCATGGGC 892
QY 3845 TGGCATGTGCACAGTGGACCTTTCCAGTGCAGCATATGCCAGATCTTTGCACGACAAA 3904
DB 892 TGCACGGCTTCGTGATCTTTTGATGCAACATGTGCGGTACACAGCAGGACCGG 952
QY 3905 TATGACTTCACACATATCCAGAGGGGCTTCA 3939
DB 952 TAGGAGTTCGTGCGACATATCCAGGAGGGGACCA 986

RESULT 8
US-08-465-590-3
Sequence 3, Application US/08465590
Patent No. 5824770
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia A.
TITLE OF INVENTION: KAROS: A T CELL PATHWAY REGULATORY GENE
NUMBER OF SEQUENCES: 164
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 STATE STREET, Suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,590
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,212
FILING DATE: 02-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/121,438
FILING DATE: 14-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,695
REFERENCE/DOCKET NUMBER: MGP-026C2DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-465-590-3

Query Match 1.3%; Score 57.4; DB 1; Length 1386;
Best Local Similarity 60.6%; Pred. No. 4.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 3; Gaps 0;
QY 3785 AATGTGTGCACCTGGCGATTGCTTTCTGGATGAAGTATGCTTTGTCATATGAGT 3844
DB 1213 AATGGCAACATGCCGGGTGCTCTTCTGGATGACGTCATGTACACCATCCATGGGC 1272
QY 3845 TGGCATGTGCACAGTGGACCTTTCCAGTGCAGCATATGCCAGATCTTTGCACGACAAA 3904
DB 1273 TGCACGGCTTCGTGATCTTTTGATGCAACATGTGCGGTACACAGCAGGACCGG 1332
QY 3905 TATGACTTCACACATATCCAGAGGGGCTTCA 3939
DB 1333 TAGGAGTTCGTGCGACATATCCAGGAGGGGACCA 1367

RESULT 9
US-08-283-300A-2
Sequence 2, Application US/08283300A
Patent No. 6172278
GENERAL INFORMATION:
APPLICANT: Georgopoulos, Katia A.
TITLE OF INVENTION: KAROS: TRANSGENIC CELLS AND ANIMALS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, Suite 510
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109

```
;
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/283,300A
; FILING DATE:
; FILING DATE:
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/238,212
; FILING DATE: 02-MAY-94
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,695
; REFERENCE/DOCKET NUMBER: MGP-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...1386
; US-08-283-300A-2

Query Match 1.3%; Score 57.4; DB 3; Length 1386;
Best Local Similarity 60.6%; Pred. No. 4.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 3785 AATGTGTGCACTGTGGCATGTCTTTCTGGATGAGTGTATGCTTTGCATATGACT 3844
Db 1213 AAGTGGCAACATGCCGGGTGCTCTTCTGGATCAGTGTATGACCATCCACATGGGC 1272
Qy 3845 TGCATGTTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGCATCTTTSCAGGACAAA 3904
Db 1273 TGCACGGCTTCCTGTATCTCTTTAGTGAACATGTGCGGTACCAACAGCCAGGACCG 1332
Qy 3905 TATGACTTCACACACATATCCAGAGGGGCTGCA 3939
Db 1333 TACGAGTTCGTGCGCACATAACCGAGGGAGCA 1367

RESULT 10
US-08-711-417C-3
; Sequence 3, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b

;
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-9906
; TELEX: 280154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1383
; US-08-711-417C-3

Query Match 1.3%; Score 57.4; DB 3; Length 1386;
Best Local Similarity 60.6%; Pred. No. 4.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 3785 AATGTGTGCACTGTGGCATGTCTTTCTGGATGAGTGTATGCTTTGCATATGACT 3844
Db 1213 AAGTGGCAACATGCCGGGTGCTCTTCTGGATCAGTGTATGACCATCCACATGGGC 1272
Qy 3845 TGCATGTTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGCATCTTTSCAGGACAAA 3904
Db 1273 TGCACGGCTTCCTGTATCTCTTTAGTGAACATGTGCGGTACCAACAGCCAGGACCG 1332
Qy 3905 TATGACTTCACACACATATCCAGAGGGGCTGCA 3939
Db 1333 TACGAGTTCGTGCGCACATAACCGAGGGAGCA 1367

RESULT 11
US-08-733-622C-16
; Sequence 16, Application US/08733622C
; Patent No. 8528634
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: AICLOS GENE
; FILE REFERENCE: 0287-03000
; CURRENT APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 60/017,646
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 60/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1386
; TYPE: cDNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1383)
; US-08-733-622C-16
```

	Query Match:	1.3%;	Score 57.4;	DB 4;	Length 1386;
	Best Local Similarity	60.8%;	Pred. No. 4.5e-07;		
Matches	99;	Conservative	0;	Mismatches	61;
				Indels	0;
				Gaps	0;
QY	3785	AAATGTGTGCATCTGGCATGTGCTTTCTCGATGAATGTATGTATGCTTTGCCATATGAGT	3844		
DB	1213	AAGTGGCAACACTGCCGGTGCTCTCTCTCGATCACGTGATGTACACCATCCACATGGGC	1272		
QY	3845	TGCCATGTGTGCATGTGGACCTTTTCCAGTGCAAGCATATSCCAGCATCTTTTCCACGGACAA	3904		
DB	1273	TGCCACGGCTTCGGTGATCTTTTGTAGTGCAACATGTGCGGCTACACACCCACGGACCG	1332		
QY	3905	TATGACTTTCACACACATATCCAGAGGGGCTGCA	3939		
DB	1333	TACGAGTTCTCGTCGCACATAACCGCAGGGGAGCA	1367		

```

RESULT 12
PCT-US95-09345-2
; Sequence 2, Application PC/TUS9509345
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS TRANSGENIC CELLS AND ANIMALS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Ascii (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09345
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/283,300
; FILING DATE: 29-JULY-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/238,212
; FILING DATE: 02-MAY-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/121,438
; FILING DATE: 14-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul J.
; REGISTRATION NUMBER: 35,695
; REFERENCE/DOCKET NUMBER: MGP-027PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
PCT-US95-09345-2

```

Matches	94	Conservative	3	Mismatches	61	Indels	0	Gaps	0
QY	3785	AAATGTGTGACATGTGGCAATGTCTTTCTGGATGAAGTATGATGCTTTGGCATATGAGT	3844						
DB	1213	AAGTGGCAACACTGGCGGGTGTCTTTCTTGGATCAGCTCATGTACACCATCCACATGGGC	1272						
QY	3845	TGCGRTGGTGCACAGTGGAGCTTTCCAGTGGCAGCATATGCCAGCATCTTTGCACGGGACAAA	3904						
DB	1273	TGCCACGGCTTCGGTGATCTCTTTTSGATGCAACATGTGCGGTACACAGCCAGGACCGG	1332						
QY	3905	TATGACTTCACAAACACATATCCAGAGGGGCGCTGCA	3939						
DB	1333	TACGAGTTTCTGCGCACATACGCGAGGGGAGCA	1367						
RESULT 13									
US-08-711-417C-165									
; Sequence 165, Application US/08711417C									
; Patent No. 622861									
; GENERAL INFORMATION:									
; APPLICANT: Geopropoulos, Katia A.									
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE									
; NUMBER OF SEQUENCES: 202									
; CORRESPONDENCE ADDRESS:									
; ADDRESSEE: Fish & Richardson P.C.									
; STREET: 225 Franklin Street									
; CITY: Boston.									
; STATE: MA									
; COUNTRY: USA									
; ZIP: 02110-2804									
; COMPUTER READABLE FORM:									
; MEDIUM TYPE: Diskette									
; COMPUTER: IBM Compatible									
; OPERATING SYSTEM: Windows 95									
; SOFTWARE: FastSeq for Windows Version 2.0b									
; CURRENT APPLICATION DATA: US/08/711.417C									
; APPLICATION NUMBER: US/08/711.417C									
; FILING DATE: 05-Sep-1996									
; PRIOR APPLICATION DATA:									
; APPLICATION NUMBER: 08/235,212									
; FILING DATE: 02-MAY-1994									
; APPLICATION NUMBER: 08/121,436									
; FILING DATE: 14-SEP-1993									
; APPLICATION NUMBER: 07/946,223									
; FILING DATE: 14-SEP-1992									
; ATTORNEY/AGENT INFORMATION:									
; NAME: Myers, Louis P.									
; REGISTRATION NUMBER: 35,965									
; REFERENCE/DOCKET NUMBER: 10									

3845 TGCCATGGTGACAGTGGAGCTTTCCAGTCGAGCATATGCCAGCATCTTTTCACGGACAAA 3904
1438 TGCCACGGCTTCGTGTATCTTTTGAGTGCACATGTGCGGTACCCACACCCAGGACCG 1497

QY
Db
QY
Db

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	1.34;	Score 57.4;	DB 5;	Length 1611;
Best Local Similarity	60.6%;	Pred. No. Se-07;		
Matches 94;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
Qy	3785	AAATGTGTGCACCTGGCATCTCTTCTGTGATGAAGTGATGTATGTTTGCATATGAGT	3844	
Db	1435	AAGTGCGAACACTCCGGTGCTCTTCTTGATCAGTCTATGACATCCACATGGCC	1494	
Qy	3845	TGCCATCGTGACAGTGGACCTTTCCAGTGCACGATATGCCAGCATCTTTGCACGGACAA	3904	
Db	1495	TGCCACGGCTTCGTGATCCTTTTGAGTGCACCATGTGGCGCTACCACGCCAGACCGG	1554	
Qy	3905	TATGACTTTCACAAACACATATCCAGAGGGGCGCTGCA	3939	
Db	1555	TACGAGTTCTCGTCGCACATAACCGCAGGGGAGCA	1589	

```

RESULT 15
US-09-435-327A-19
; Sequence 19: Application US/09435327A
; Patent No. 6537766
; GENERAL INFORMATION:
; APPLICANT: Uckun, Fatih M.
; APPLICANT: Crotty, Mya L.
; TITLE OF INVENTION: IKAROS ISOFORMS AND MUTANTS
; FILE REFERENCE: 12152.35U501
; CURRENT APPLICATION NUMBER: US/09/435,327A
; CURRENT FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/107,229
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 27

```

```
; SOFTWARE: PatenIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-435-327A-19

Query Match:          1.3%; Score 57.4; DB 4; Length 1788;
Best Local Similarity 60.6%; Pred.No.5.4e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0; O:

QY      3785 AAATGTGTCACATGTGCATGTCTTTCTGGATGAAGTGATGTATGCTTTGCCATATCAGT 3844
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1615 AAGTGGCAACACTGCCGGGTGCTCTTCTCGATCAGCTCATGTACACATCCACATGGC 1674
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      3845 TGCCATGGTGACAGTGGACGCTTTCCAGTGACGACATATGCCAGCATCTTTTCACCGACAAA 3904
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1675 TGCACGGCTTCGGTGATCCTTTTGAGTGCMAAGATGTSCGGCTACCACACCGGACCGG 1734
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      3905 TATGACTTCACAACACATATCCAGAGGGGCGCTCCA 3939
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1735 TAGCAGTCTCGTCGGCATAAACCGCGGGSGAGCA 1769
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 29, 2003, 15:29:16
Job time : 237 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 29, 2003, 11:49:17 ; Search time 1125 Seconds

(without alignments)
10881.588 Million cell updates/sec

Title: US-09-702-216-1_COPY_5000

Perfect score: 4501

Sequence: 1 atgtcttctgtcttctga.....agaaaaatggcgccctctca 450:

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1811591 seqs, 1359896230 residues

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubna/PCr_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubna/PCrUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	4501	100.0	10011	14	US-10-177-293-471
2	4501	100.0	10102	14	US-10-198-846-10440
3	1060.4	23.6	1062	12	US-10-029-386-25248
4	542.2	12.0	879	14	US-10-198-846-3873
5	510.4	11.3	512	12	US-10-029-386-11541
6	453.4	10.1	510	14	US-10-198-846-3994
7	398.8	8.9	520	14	US-10-198-846-8958
8	389.2	8.6	836	14	US-10-198-846-3330
9	324.4	7.2	427	14	US-10-198-846-5878
10	322.8	7.2	476	11	US-09-918-995-10414
11	275.4	6.1	305	9	US-09-778-320-110
12	275.4	6.1	305	9	US-09-910-689-110
13	275.4	6.1	305	13	US-10-010-742-110
14	210	4.7	810	14	US-10-198-846-7424
15	133.4	3.0	594	14	US-10-198-846-9634
16	119.2	2.6	467	14	US-10-198-846-419

17	99.4	2.2	968	14	US-10-198-846-7398
18	80.6	1.8	166	10	US-09-728-444-1065
19	65.4	1.5	2177	14	US-10-084-917-291
20	65.4	1.5	2610	14	US-10-177-293-173
21	65	1.4	569	14	US-10-198-846-11151
22	65	1.4	1026	14	US-10-198-846-10906
23	65	1.4	1475	10	US-09-964-824A-293
24	65	1.4	2025	15	US-10-081-327-46
25	65	1.4	2365	12	US-10-097-926A-78
26	65	1.4	2673	14	US-10-177-293-178
27	65	1.4	2709	14	US-10-177-293-177
28	65	1.4	3029	14	US-10-084-817-331
29	65	1.4	3118	14	US-10-177-293-175
30	65	1.4	318	14	US-10-198-846-11426
31	64.6	1.4	514	10	US-09-764-864-196
32	64.6	1.4	523	10	US-09-764-864-1623
33	64.6	1.4	523	10	US-09-764-864-1624
34	64.6	1.4	523	10	US-09-764-864-1625
35	63.6	1.4	674	14	US-10-198-846-7898
36	63.4	1.4	1326	10	US-09-749-728B-12
37	63.4	1.4	3411	14	US-10-198-846-9872
38	62.8	1.4	3671778	12	US-10-312-841-2
39	62	1.4	682	14	US-10-198-846-2403
40	57.4	1.3	1024	10	US-09-755-830-7
41	57.4	1.3	1386	10	US-09-755-830-2
42	56.6	1.3	2453	14	US-10-198-846-13080
43	55.8	1.2	1498	12	US-10-007-926A-4
44	55.6	1.2	856	14	US-10-198-846-4034
45	53.3	1.2	1647	14	US-10-106-698-1074

ALIGNMENTS

RESULT 1

US-10-177-293-471

Sequence 471, Application: US/10177293

Publication No. US2003012426A1

GENERAL INFORMATION:

APPLICANT: Lillic, James

APPLICANT: Gatt, Karen

APPLICANT: Zhao, Xumel

APPLICANT: Gannavarpu, Manjula

APPLICANT: Kamatkar, Shubhang:

APPLICANT: Nertens, Maureen

APPLICANT: Myer, Vic

APPLICANT: Wang, Youzhen

APPLICANT: Xu, Yongyao

APPLICANT: Hoersch, Sebastian

APPLICANT: Monahan, John

APPLICANT: Meyers, Rachel E.

APPLICANT: Bast Jr., Robert C.

APPLICANT: Hortobagyi, Gabriel N.

APPLICANT: Puztai, La'os

APPLICANT: Meric, Funda

APPLICANT: Sahin, Aysegul

APPLICANT: Mills, Gordon B.

TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,

FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER

FILE REFERENCE: MRI-038

CURRENT APPLICATION NUMBER: US/10/177,293

CURRENT FILING DATE: 2002-06-21

PRIOR APPLICATION NUMBER: US 60/299,887

PRIOR FILING DATE: 2001-06-21

PRIOR APPLICATION NUMBER: US 60/301,572

PRIOR FILING DATE: 2001-06-27

PRIOR APPLICATION NUMBER: US 60/306,501

PRIOR FILING DATE: 2001-07-18

PRIOR APPLICATION NUMBER: US 60/325,002

PRIOR FILING DATE: 2001-09-25

PRIOR APPLICATION NUMBER: US 60/362,595

PRIOR FILING DATE: 2002-03-05

PRIOR APPLICATION NUMBER: US 60/xxx,xxx

2480	DB		TCAGTGCCTACAGTGTTCATTCACCAACCCCTGAGTGTACTCCTCTTTTCACTATGA	2539
2041	QY		AAAGTGTGCATGAGTCCCAAGCATCGCATGTCAAAACAAGAGCAAAATCACTCTGCAAGGATC	2100
2540	DB		AAAGTGTGCATGAGTCCCAAGCATCGCATGTCAAAACAAGAGCAAAATCACTCTGCAAGGATC	2599
2101	QY		GGATGGCGAGCATCTGTTCMAAGGAAAGCAAGAACACTCATGTACCAATGTGATTTTAT	2160
2600	DB		GGATGGCGAGCATCTGTTCMAAGGAAAGCAAGAACACTCATGTACCAAAATGTGATTTTAT	2659
2161	QY		TACCAAGTGGAAAGAGAGATTTCCCGACACTACAGAGAGCACACAGTGTGTACAAATG	2220
2660	DB		TACCAAGTGGAAAGAGAGATTTCCCGACACTACAGAGAGCACACAGTGTGTACAAATG	2719
2221	QY		CGCTCAGTGCAGTTTACAGCTCCGATACTCAGTCACTACTGGAGCACTTCAACACTGT	2280
2720	DB		CGCTCAGTGCAGTTTACAGCTCCGATACTCAGTCACTACTGGAGCACTTCAACACTGT	2779
2281	QY		TCACTGCCAGGACAGGACATCACTACAGCCACCGCGAGAGGAGCGGTATGCCATATC	2340
2780	DB		TCACTGCCAGGAAACAGGACATCACTACAGCCACCGCGAGAGGAGCGGTATGCCATATC	2839
2341	QY		CACCATCAAAAGAGAGGCCAAAATTGACTTTCAGGGTCTACAAATCTGTAACTCCAGACTC	2400
2840	DB		CACCATCAAAAGAGAGGCCAAAATTGACTTTCAGGGTCTACAAATCTGTAACTCCAGACTC	2899
2401	QY		TAAAAATGGAGAGCGCAGTTTCTCAGAGTGTGTTGAAGAGAGAGAACTGGAAGAGAGGA	2460
2900	DB		TAAAAATGGAGAGCGCAGTTTCTCAGAGTGTGTTGAAGAGAGAGAACTGGAAGAGAGGA	2959
2461	QY		CGGGCTCAAAAGAGAAAGTTTGGACCCGAGAGTTCAGTGATGACTTCCTGCGAAATGTGACTTG	2520
2960	DB		CGGGCTCAAAAGAGAAAGTTTGGACCCGAGAGTTCAGTGATGACTTCCTGCGAAATGTGACTTG	3019
2521	QY		GAGAGGGGCAGACATCTCTCGGGGGAGTCCGTATACACCAAGCAAGCTCGGGGTGCT	2580
3020	DB		GAGAGGGGCAGACATCTCTCGGGGGAGTCCGTATACACCAAGCAAGCTCGGGGTGCT	3079
2581	QY		GAGCGCTGTCTGGCACCCAGAGCAGACAAAGACTCTAAGGATAGTCCCAATGTGGA	2640
3080	DB		GAGCGCTGTCTGGCACCCAGAGCAGACAAAGACTCTAAGGATAGTCCCAATGTGGA	3139
2641	QY		GGCGCGCCCATCTGGCGGCAGCTATTTATGGCTTGGCTGTGGAAACCAAGGATTCCTGCA	2700
3140	DB		GGCGCGCCCATCTGGCGGCAGCTATTTATGGCTTGGCTGTGGAAACCAAGGATTCCTGCA	3199
2701	QY		GGGGGGCGCAGCTGGCGAGAGAGTCTGGGGCCCTCCCCACAGCATCTCTCATCTGGG	2760
3200	DB		GGGGGGCGCAGCTGGCGAGAGAGTCTGGGGCCCTCCCCACAGCATCTCTCATCTGGG	3259
2761	QY		AGAAAACAACTCCAAGGATGAATCCCAAGTCCCTGTTTACGGAGCGTAGAGGCTCCGCTGT	2820
3260	DB		AGAAAACAACTCCAAGGATGAATCCCAAGTCCCTGTTTACGGAGCGTAGAGGCTCCGCTGT	3319
2821	QY		TTTTTGTGCCAATTGCCCTGACCAACAAGACCTCTCTCTGGCGAAGAAATGCAAAATGGCGG	2880
3320	DB		TTTTTGTGCCAATTGCCCTGACCAACAAGACCTCTCTCTGGCGAAGAAATGCAAAATGGCGG	3379
2881	QY		ATATGTATGCAACCGCTGTGCCTCTTACCAGAGCTTCACTCGACTCCCAAGGCTTTAAA	2940
3380	DB		ATATGTATGCAACCGCTGTGCCTCTTACCAGAGCTTCACTCGACTCCCAAGGCTTTAAA	3439
2941	QY		CATCATTTAAACAAAACAAACGGTGTAGCAGATTTATTAGGAGGAGAAACAAGAAGCGCTTAA	3000
3440	DB		CATCATTTAAACAAAACAAACGGTGTAGCAGATTTATTAGGAGGAGAAACAAGAAGCGCTTAA	3499
3001	QY		CCCAGAGGCATCTCAGGCTGAGCAGCTCAACAAACAGCAGAGGGCGCAGCAATGAGGAGCA	3060
3500	DB		CCCAGAGGCATCTCAGGCTGAGCAGCTCAACAAACAGCAGAGGGCGCAGCAATGAGGAGCA	3559
3061	QY		AGTCAATGGAAAGCCCGTTTAGAGAGGAGGTTCAGAAAGATCATCTTAATCTGAAAGTCAACAG	3120

Dp	3560	AGTCAATGGAAGCCCGTTAGAGGAGGAGCTAGAGAGATCATCTTAACTGAAAGTCAACAGAG	3619
Qy	3121	AGAAAATCCCACTCCCAGAGCTAAGTAAATACGAAGGCCAGGGTTCATTGACTAAAAAGCCA	3180
Dp	3620	AGAAATCCCACTCCCAGAGCTAAGTAAATACGAAGGCCAGGGTTCATTGACTAAAAAGCCA	3679
Qy	3181	TTCTGCTCAGCAGCCAGTCTCGTGTAGCCCAAACTCTGATATTCACAAAAGGATGCAACC	3240
Dp	3680	TTCTGCTCAGCAGCCAGTCTCGTGTAGCCCAAACTCTGATATTCACAAAAGGATGCAACC	3739
Qy	3241	TTTGCAATTCAGATAAAAAGTCCCTCAGGAAAGTACTCGAGATCCAGGAATAAGTTCATC	3300
Dp	3740	TTTGCAATTCAGATAAAAAGTCCCTCAGGAAAGTACTCGAGATCCAGGAATAAGTTCATC	3799
Qy	3301	CGTATCTGAAGGGAGGAAGTCTCGAGAGGCGAGTCTCTATAGAAAAGTACATGAGAAC	3360
Dp	3800	CGTATCTGAAGGGAGGAAGTCTCGAGAGGCGAGTCTCTATAGAAAAGTACATGAGAAC	3859
Qy	3361	TGCGAAACCCAAATATTACACACAGCAGCCCTATTGAAAAGTACCAGTACCCCACT	3420
Dp	3860	TGCGAAACCCCAAAATTATCACACAGCAGCCCTATTGAAAAGTACCAGTACCCCACT	3919
Qy	3421	TTTTGGACTTCCCTTTGTACATTAATGACTTCCAGAGTGAAGCTGATTGGCTCGGGTTCTG	3480
Dp	3920	TTTTGGACTTCCCTTTGTACATTAATGACTTCCAGAGTGAAGCTGATTGGCTCGGGTTCTG	3979
Qy	3481	GAGTAATAATAGCTCTCCGTTCTCGGATCCGCACTACTCTGAGTACACCTGCTGCCT	3540
Dp	3980	GAGTAATAATAGCTCTCCGTTCTCGGAAATCCGCACTACTCTGAGTACACCTGCTGCCT	4039
Qy	3541	ACCAAAATCCCTTGCCAAAATATATGCGCTTATCCCACTTCAATCTGCTCTCATATTTTC	3600
Dp	4040	ACCAAAATCCCTTGCCAAAATATATGCGCTTATCCCACTTCAATCTGCTCTCATATTTTC	4099
Qy	3601	AGCTGTGGATCAGACATGACATTCCTCTAGATTTTGCGCATCAAGCATTTCCAGACCTGG	3660
Dp	4100	AGCTGTGGATCAGACAAATGACATTCCTCTAGATTTTGCGCATCAAGCATTTCCAGACCTGG	4159
Qy	3661	GCMAATCGCAAAACGGTGCCTTCCAAGGAGAAAACGAAAGGCACCACTAATGTAAAAAATGA	3720
Dp	4160	GCMAATCGCAAAACGGTGCCTTCCAAGGAGAAAACGAAAGGCACCACTAATGTAAAAAATGA	4219
Qy	3721	AGTCCCTTTGAATGTAGTATAAAAACAGAGAAAGTTGATAGAAAGTACTCAAGATGAACCTTC	3780
Dp	4220	AGTCCCTTTGAATGTAGTATAAAAACAGAGAAAGTTGATAGAAAGTACTCAAGATGAACCTTC	4279
Qy	3781	AACAAAATGTGCACGTGCGCATTTCTTCTGAGATGAAGTATGATGCTTTGCAATAT	3840
Dp	4280	AACAAAATGTGCACATGTGCGCATTTCTTCTGAGATGAAGTATGATGCTTTGCAATAT	4339
Qy	3841	GAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATGCGAGCATCTTTGCACGGA	3900
Dp	4340	GAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATGCGAGCATCTTTGCACGGA	4399
Qy	3901	CAAAATATGACTTCCACACACATATCCAGAGGGCCCTGCATAGGAACCAATGCACAAGTGGGA	3960
Dp	4400	CAAAATATGACTTCCACACACATATCCAGAGGGCCCTGCATAGGAACCAATGCACAAGTGGGA	4459
Qy	3961	AAAAAATGGAAAACCTTAAGAGTAAAAACCTTTAGCACTTAGCACAAATTAATAGAAATAGG	4020
Dp	4460	AAAAAATGGAAAACCTTAAGAGTAAAAACCTTTAGCACTTAGCACAAATTAATAGAAATAGG	4519
Qy	4021	TTTTCTTGATGGGAATTCATAGCTTGTAATGCTTATGAAGCACTTATTAATAAATACT	4080
Dp	4520	TTTTCTTGATGGGAATTCATAGCTTGTAATGCTTATGAAGCACTTATTAATAAATACT	4579
Qy	4081	TCAATAGAGCTGCTTATCCAAACATGAAATTCCTCTTTTGTATTCTCTTTCTTTTGATG	4140
Dp	4580	TCATAGAGCTGCTTATCCAAATGAAATTCCTCTTTTGTATTCTCTTTCTTTTGATG	4639
Qy	4141	AGTAGGTACCAAGATTAAAAAGTGAGATAAATGGTCAATGAGAAAGATGGAGAATGGT	4200
Dp	4640	AGTAGGTACCAAGATTAAAAAGTGAGATAAATGGTCAATGAGAAAGATGGAGAATGGT	4699

Db 1640 CTGTAAATTCCTGCAATTTCACTTATATGGCACTCATCCACGAAATAGAACACACATTT 1699
Qy 1201 TCTTCAGACTCACCCAAACAAAATAAAGCTTCTCTCCCTCCTCTGAGGTTGCAAAACC 1260
Db 1700 TCTTCAGACTCACCCAAACAAAATAAAGCTTCTCTCCCTCCTCTGAGGTTGCAAAACC 1759
Qy 1261 TTCCAGAGAAAATCTTAACAGTCCATCCCTGACCTTCAATCCAGTGAATCTGGAGACTT 1320
Db 1760 TTCCAGAGAAAATCTTAACAGTCCATCCCTGACCTTCAATCCAGTGAATCTGGAGACTT 1819
Qy 1321 GGGAAAATGGCAGGACAAAGATAACAGTCAAGCAGGAGATGACACTCTCTGTTGGTACTC 1380
Db 1920 GGGAAAATGGCAGGACAAAGATAACAGTCAAGCAGGAGATGACACTCTCTGTTGGTACTC 1879
Qy 1381 AGTGCCCATTAAGGCCCTCGATTCTCTAGACAAAATGGTACAGAGCCACCACTTACTA 1440
Db 1880 AGTGCCCATTAAGGCCCTCGATTCTCTAGACAAAATGGTACAGAGCCACCACTTACTA 1939
Qy 1441 CTGGTGTAAATTTGTAGTTTCAGCTGTGAGTCACTAGCTCACTTAAACTCTGAACA 1500
Db 1940 CTGGTGTAAATTTGTAGTTTCAGCTGTGAGTCACTAGCTCACTTAAACTCTGAACA 1999
Qy 1501 TTATGGCAAGCAGCAGGACAGTGCAGTCAGCGGCCCTTAATCCAGACTTAAATGATAA 1560
Db 2000 TTATGGCAAGCAGCAGGACAGTGCAGTCAGCGGCCCTTAATCCAGACTTAAATGATAA 2059
Qy 1561 GCTTTCCAGGGGCTCTGTCAATTAATCAGATGATCTAGCCAAAGTTCCAGAGGAGAC 1620
Db 2060 GCTTTCCAGGGGCTCTGTCAATTAATCAGATGATCTAGCCAAAGTTCCAGAGGAGAC 2119
Qy 1621 AATGACCAAGACAGCAAGAGCTCGAGTGGGGTAAAAAGAGAGACTTCTCCAGCAAGG 1680
Db 2120 AATGACCAAGACAGCAAGAGCTCGAGTGGGGTAAAAAGAGAGACTTCTCCAGCAAGG 2179
Qy 1681 AGCCGAGGATAATATGTTAAGAGCTATAATTTGTAGTTCTGTGACTTCGGATATCCAA 1740
Db 2180 AGCCGAGGATAATATGTTAAGAGCTATAATTTGTAGTTCTGTGACTTCGGATATCCAA 2239
Qy 1741 AAGCCATGGCCCTGATGATTAATTTAGTGGGGCCACTTCTCCGTCATTATCAACAGTCCA 1800
Db 2240 AAGCCATGGCCCTGATGATTAATTTAGTGGGGCCACTTCTCCGTCATTATCAACAGTCCA 2299
Qy 1801 TAACATTCACAAAGTGTACCAATTAACACTGTCCATCTGTGCCAGAGGACTTTGACGCC 1860
Db 2300 TAACATTCACAAAGTGTACCAATTAACACTGTCCATCTGTGCCAGAGGACTTTGACGCC 2359
Qy 1861 AGAAAGCACTTTGGAGAAATTAATTCGTTTGTGTTGTTAGTAAAGTAAATTTGTTCCA 1920
Db 2360 AGAAAGCACTTTGGAGAAATTAATTCGTTTGTGTTGTTAGTAAAGTAAATTTGTTCCA 2419
Qy 1921 CTGTGCACTCTTGTCTGCACTTGTCTCCTGGGGGGCTGGAAGCTCGCGAGTCAAAACA 1980
Db 2420 CTGTGCACTCTTGTCTGCACTTGTCTCCTGGGGGGCTGGAAGCTCGCGAGTCAAAACA 2479
Qy 1981 TCAGTGCCATCAGTTGTTCAATTCACCAACCCTGAGCTAGATGACTCTCTTTCATGATCA 2040
Db 2480 TCAGTGCCATCAGTTGTTCAATTCACCAACCCTGAGCTAGATGACTCTCTTTCATGATCA 2539
Qy 2041 AAGTGTGCATGAGTCCCAAGCATCGGATGTCAAAACAAAGAACAAATCACCTGCAAGGATC 2100
Db 2540 AAGTGTGCATGAGTCCCAAGCATCGGATGTCAAAACAAAGAACAAATCACCTGCAAGGATC 2599
Qy 2101 GGATGGGAGCAGTCTGTCAAGGAAAGCAAGAACACTCATGTACCAAAATGTGATTTTC 2160
Db 2600 GGATGGGAGCAGTCTGTCAAGGAAAGCAAGAACACTCATGTACCAAAATGTGATTTTC 2659
Qy 2161 TACCAGTGTGAAGAGAGATTTCCGACACTACAGAGAGCAGACAGCTGCTACAAATG 2220
Db 2660 TACCAGTGTGAAGAGAGATTTCCGACACTACAGAGAGCAGACAGCTGCTACAAATG 2719
Qy 2221 CCGTCAGTGCAGTTTACAGCTGCCGATACTCAGTCACTACTGGAGCACTTCAACACTGT 2280
Db 2720 CCGTCAGTGCAGTTTACAGCTGCCGATACTCAGTCACTACTGGAGCACTTCAACACTGT 2779

Qy 2281 TCACCTCCAGGAACAGGACATCACTACAGCCAAACGGCGAAGAGGAGGTCTATGCCATATC 2340
Db 2780 TCACCTCCAGGAACAGGACATCACTACAGCCAAACGGCGAAGAGGAGGTCTATGCCATATC 2839
Qy 2341 CACCATCAAGAGAGGCCCAAAATGACITTCAGGTCTACATCTGTAACCTCCAGACTC 2400
Db 2840 CACCATCAAGAGAGGCCCAAAATGACITTCAGGTCTACATCTGTAACCTCCAGACTC 2899
Qy 2401 TAAAAATGGGAGAGCCAGTTTCTGAGAGTGTGTGAAGAGAGAGAGCTGGAAAGAAAG 2460
Db 2900 TAAAAATGGGAGAGCCAGTTTCTGAGAGTGTGTGAAGAGAGAGAGCTGGAAAGAAAG 2959
Qy 2461 CCGGCTCAAGAGAGAAATTTGGACCGAGAGTTCCAGTGAAGACCTTCGAAATGTGACTTG 2520
Db 2960 CCGGCTCAAGAGAGAAATTTGGACCGAGAGTTCCAGTGAAGACCTTCGAAATGTGACTTG 3019
Qy 2521 GAGAGGGGAGACATCCTCGCGGAGTCCGTCAATACACCCAAAGCAAGCTCGGCTGTCT 2580
Db 3020 GAGAGGGGAGACATCCTCGCGGAGTCCGTCAATACACCCAAAGCAAGCTCGGCTGTCT 3079
Qy 2581 GAGCCTGTGTCTGGCACTCAAGAGAGAGACAAAGACTCTAAGGATAGTCCCAATGTGA 2640
Db 3080 GAGCCTGTGTCTGGCACTCAAGAGAGAGACAAAGACTCTAAGGATAGTCCCAATGTGA 3139
Qy 2641 GCGCGGCCATCTGGCGGACCTATTTATGGCTTGGCTGTGGAAACCAAGGATTCCTGCA 2700
Db 3140 GCGCGGCCATCTGGCGGACCTATTTATGGCTTGGCTGTGGAAACCAAGGATTCCTGCA 3199
Qy 2701 GCGCGGCCAGCTGGCGGAGAGAGTCTGGGCGCTCTCCCGAGCAGTATCTTGCACTCGG 2760
Db 3200 GCGCGGCCAGCTGGCGGAGAGAGTCTGGGCGCTCTCCCGAGCAGTATCTTGCACTCGG 3259
Qy 2761 AGAAACAACTCCAAAGATGAATCCCACTCTCTGTTACGAGGCGTGAAGCTCCGGTGT 2820
Db 3260 AGAAACAACTCCAAAGATGAATCCCACTCTCTGTTACGAGGCGTGAAGCTCCGGTGT 3319
Qy 2821 TTTTGTGCCAATTTGCTGACCAAAAGACCTCTCTCTGGCAAAAGATTCGCAATGGCGG 2880
Db 3320 TTTTGTGCCAATTTGCTGACCAAAAGACCTCTCTCTGGCAAAAGATTCGCAATGGCGG 3379
Qy 2881 ATATGTATGCAAGCGCTGTGGCTCTACAGAGCTTCACTCCAGGCTTTTAA 2940
Db 3380 ATATGTATGCAAGCGCTGTGGCTCTACAGAGCTTCACTCCAGGCTTTTAA 3439
Qy 2941 CATCATTTAAACAAACAAAGCTGAGCAGATTTATAGAGGAGAACAAAGAGCGCTTAA 3000
Db 3440 CATCATTTAAACAAACAAAGCTGAGCAGATTTATAGAGGAGAACAAAGAGCGCTTAA 3499
Qy 3001 CCGAGAGGCACTTCAGGCTGAGCAGTCAACAAACAGCAGAGGGGCGAGCAATGAGAGCA 3060
Db 3500 CCGAGAGGCACTTCAGGCTGAGCAGTCAACAAACAGCAGAGGGGCGAGCAATGAGAGCA 3559
Qy 3061 AGTCATTTGAGAGCGCTTACAGAGGAGGTCAAGATCATCTAACTGAAGTCAACAGAG 3120
Db 3560 AGTCATTTGAGAGCGCTTACAGAGGAGGTCAAGATCATCTAACTGAAGTCAACAGAG 3619
Qy 3121 AGAAATTCACACTCCCAGGCTTAAGTAAATACGAAGCCCAAGGCTTCACTGACTTAAAGCA 3180
Db 3620 AGAAATTCACACTCCCAGGCTTAAGTAAATACGAAGCCCAAGGCTTCACTGACTTAAAGCA 3679
Qy 3181 TTCTGCTCAGCAGCCAGTCTGTGTGAGCAAACTCTGTGATATTCAGAAAGATGCAACC 3240
Db 3680 TTCTGCTCAGCAGCCAGTCTGTGTGAGCAAACTCTGTGATATTCAGAAAGATGCAACC 3739
Qy 3241 TTTCGCACTTCAGATTAAGTCTCTCAGGAAAGTCTGAGATCCAGGAAATAGTTCATC 3300
Db 3740 TTTCGCACTTCAGATTAAGTCTCTCAGGAAAGTCTGAGATCCAGGAAATAGTTCATC 3799
Qy 3301 CGTACTCTGAAGGAAAGAGGATTTCTCAGAGAGGAGTCTCTATAGAAAGTACATGAGACC 3360
Db 3800 CGTACTCTGAAGGAAAGAGGATTTCTCAGAGAGGAGTCTCTATAGAAAGTACATGAGACC 3859

```
Db 512 CAGAGGGGAGCAATGAGGAGCAAGTCAATGAGGCCGCTTATGAGGAGGCTCAGAGAT 453
Qy 3098 CATCTTAACCTGAAGTCAACAGAGAGAAATCCACTCCCCAGCCTAAGTAAATACGAAGCC 3157
Db 452 CATCTAACTGAAGTCAACAGAGAGAAATCCACTCCCCAGCCTAAGTAAATACGAAGCC 393
Qy 3158 CAGGGTTCATTGACTAAAGCCATTCTGCTCAGAGCCAGCTCTGCTCAGCCAACTCTG 3217
Db 392 CAGGGTTCATTGACTAAAGCCATTCTGCTCAGAGCCAGCTCTGCTCAGCCAACTCTG 333
Qy 3218 GATATTCACAAAAGGATGCAACCTTTGCAATTCAGATAAAAAGTCTCTCAGGAAGTACT 3277
Db 332 GATATTCACAAAAGGATGCAACCTTTGCAATTCAGATAAAAAGTCTCTCAGGAAGTACT 273
Qy 3278 GGAGATCCAGGAATAGTTCATCCGTATCTGAAGGGAAGAAAGTTCTGAGAGAGCAAT 3337
Db 272 GGAGATCCAGGAATAGTTCATCCGTATCTGAAGGGAAGAAAGTTCTGAGAGAGCAAT 213
Qy 3338 CCTATAGAAAAGTACATGAGACCTGCGAATCCCAATTAATTCACCAAGGAGCCCT 3397
Db 212 CCTATAGAAAAGTACATGAGACCTGCGAATCCCAATTAATTCACCAAGGAGCCCT 153
Qy 3398 ATTGAAAAGTACCAGTACCACCTTTTGGACTTCCCTTTGCTACATATGACTTCCAGAGT 3457
Db 152 ATTGAAAAGTACCAGTACCACCTTTTGGACTTCCCTTTGCTACATATGACTTCCAGAGT 93
Qy 3458 GAAGCTGATTGGCTGGGTTCTGGAGTAAATATAAGCTCCCTTCTGGGAATCCGCAC 3517
Db 92 GAAGCTGATTGGCTGGGTTCTGGAGTAAATATAAGCTCTCCCTTCTGGGAATCCGCAC 33
Qy 3518 TACTTGAGTCACGTCGCTGCGCTACCAATCC 3549
Db 32 TACTTGAGTCACGTCGCTGCGCTACCAATCC 1
```

RESULT 6

```
US-10-198-846-3994
; Sequence 3994, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3994
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-3994
```

```
Query Match
Best Local Similarity 10.1%; Score 453.4; DB 14; Length 510;
Matches 454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2139 CATGTACCAATGTGATTTTATTACCAAGTGAAGAGAGATTTCCCGACACTACAGGA 2198
Db 56 CAGGTACCAATGTGATTTTATTACCAAGTGAAGAGAGATTTCCCGACACTACAGGA 115
Qy 2199 GAGCACACAGTCTGCTACAAATGCGGTGAGTGCAGTTTACAGTGGCGGATACGTAC 2258
Db 116 GAGCACACAGTCTGCTACAAATGCGGTGAGTGCAGTTTACAGTGGCGGATACGTAC 175
```

```
Qy 2259 TACTGGAGCACTTCAACACACTGTTTCACTGCCAGGAACAGGACATCACTACAGCCACGGCG 2318
Db 176 TACTGGAGCACTTCAACACACTGTTTCACTGCCAGGAACAGGACATCACTACAGCCACGGCG 235
Qy 2319 AAGAGGACGGTTCATGCCATATCCACCATCAAGAGAGGAGCCCAAAATTGACTTCAGGGTCT 2378
Db 236 AAGAGGACGGTTCATGCCATATCCACCATCAAGAGAGGAGCCCAAAATTGACTTCAGGGTCT 295
Qy 2379 ACAATCTGCTTAACCTCCAGACTCTTAAATGGAGAGCCAGTTTCTGAAGTGTGTGAAGA 2438
Db 296 ACAATCTGCTTAACCTCCAGACTCTTAAATGGAGAGCCAGTTTCTGAAGTGTGTGAAGA 355
Qy 2439 GAGAGAGCTGGAGAGAGAGACCGGCTCAAGAGAGAAAGTTTGACCGAGAGTTCAGTG 2498
Db 356 GAGAGAGCTGGAGAGAGAGACCGGCTCAAGAGAGAAAGTTTGACCGAGAGTTCAGTG 415
Qy 2499 ATGACCTTCGCAATGTGACTTTGGAGAGGGGACACATCTCTCGGGGGAGTCCGTACACA 2558
Db 416 ATGACCTTCGCAATGTGACTTTGGAGAGGGGACACATCTCTCGGGGGAGTCCGTACACA 475
Qy 2559 CCCAAGCAAGCTCGGGCTGCTGACGGCTGTGTCT 2593
Db 476 CCCAAGCAAGCTCGGGCTGCTGACGGCTGTGTCT 510
```

RESULT 7

```
US-10-198-846-8958
; Sequence 8958, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8958
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2..5
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8958
```

```
Query Match
Best Local Similarity 8.9%; Score 398.8; DB 14; Length 520;
Matches 400; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1418 GGTACAGAGGCCACAGTTACTACTGGTGAAATTTTGTAGTTTACGTGTGAGTCACT 1477
Db 111 GGTACAGAGGCCACAGTTACTACTGGTGAAATTTTGTAGTTTACGTGTGAGTCACT 170
Qy 1478 AGCTCACTTAACCTGCTAGACATATGCAAGCAGCAGCAGCAGTCCAGTCCAGCGGC 1537
Db 171 AGCTCACTTAACCTGCTAGACATATGCAAGCAGCAGCAGTCCAGTCCAGCGGC 230
Qy 1538 CTTAATCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCACTAATCAGAAATGACTA 1597
Db 231 CTTAATCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCACTAATCAGAAATGACTA 290
Qy 1598 CCCAAAAGTTCCAGAGGAGAGACAAATGACCAAGACAGACAAAGAGCTCAGTGGGCTAAA 1657
Db 291 CCCAAAAGTTCCAGAGGAGAGACAAATGACCAAGACAGACAAAGAGCTCAGTGGGCTAAA 350
```

QY	1658	AAGAAGGACTTCTCCAGCAAGGAGCCGAGGATAATATGGTAACGAGCTATAATTGTCCAG	1717
Db	351	AAGAAGGACTTCTCTAGCAAGGACCGAGGATAATATGGTAACGAGCTATAATTGTCCAG	410
QY	1718	TTCTGTGACTTCCGATATTTCCAAAGGCCATGGGCCCTGATGTAATTGTAGTGGGGCCACTT	1777
Db	411	TTCTGTGACTTCCGATATTTCCAAAGGCCATGGGCCCTGATGTAATTGTAGTGGGGCCACTT	470
QY	1778	CTCCGCTCAATTATCAACAGCTCCATAACATTTACAAGGTACC	1819
Db	471	CTCCGCTCAATTATCAACAGCTCCATAACATTTACAAGGTACC	512

RESULT 8

```

US-10-198-846-3330/c
; Sequence 3330, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Yuzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-043
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3330
; LENGTH: 936

```

```

Query Match      8.6%; Score 389.2; DB 14; Length 836;
Best Local Similarity 96.5%; Pred.No.1.7e-101;
Matches 409; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 1396 CCTGATCTCCCTAGACAAATGGTACAGAGCCACCAGTTACTACTGTGTAAATTTTG 1455
DB 486 CCACTAGTTCTAGAGCGCCGAGGTACAGAGGCCACCAGTTACTACTGGTGTAAATTTTG 427

QY 1456 TAGTTTCAGCTGTGAGTGCATCTTAGTTCACCTTAAACTGTGTAGACATTTAGCAAGACGA 1515
DB 426 TAGTTTCAGCTGTGAGTGCATCTTAGTTCACCTTAAACTGTGTAGAACATTTATGGCAAGACGA 367

QY 1516 CGAGCAGTCGAGTCAGCGCGCCTTAATCCAGAGTTAATGATTAAGCTTTCCAGCGGCTC 1575
DB 366 CGAGCAGTCGAGTCAGCGCGCCTTAATCCAGAGTTAATGATTAAGCTTTCCAGCGGCTC 307

QY 1575 TGTCATTAATCAGAAATGATCAGCCAAAAGTTTCAAGSAGAGACAATGACCAAGACAGA 1635
DB 306 TGTCATTAATCAGAAATGATCAGCCAAAAGTTTCAAGSAGAGACAATGACCAAGACAGA 247

QY 1636 CAGAGCTCGAGTGGGCTTAAGAAGSAGACTTCTCCAGCAAGSAGAGCGGAGGATTAATAT 1695
DB 246 CAGAGCTCGAGTGGGCTTAAGAAGSAGACTTCTCCAGCAAGSAGAGCGGAGGATTAATAT 187

QY 1696 GGTAAACGAGCTATAATGTGTCAGTTCTGTGACTTCGATATTCCAAAGCCACTGSCCTGA 1755
DB 196 GG-ACGAGCTATAATGTGTCAGTTCTGTGACTTCGATATTCCAAAGCCACTGSCCTGA 129

QY 1756 TGTAACTTAGTGGGCGACTTTCGTCATTTATCAAGAGCTCCTATACATTCACAGTG 1815
DB 128 TGTAACTTAGTGGGCGACTTTCGTCATTTATCAAGAGCTCCTATACATTCACAGTG 69

QY 1816 TACC 18-9
DB 68 TACC 65

RESULT 9
US-10-198-846-5878
; Sequence 5878, Application US/10198846
; Publication No. US2003009974N1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 66/336,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14284
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5878
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 104, 128, 332, 145, 158, 167, 173, 177, 207, 210, 250, 254,
; LOCAT-CN: 236, 308, 324, 369, 394, 398, 407
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-5878

Query Match      7.2%; Score 324.4; DB 14; Length 427;
Best Local Similarity 91.7%; Pred.No.6.1e-83;
Matches 354; Conservative 0; Mismatches 30; Indels 3; Gaps 3;

QY 1418 GGTACAGAGGCCACACAGTACTACTGTTGTAAATTTGTAGTTTCAGTGTGAGTCATCT 1477

```



```
Db 27 GGTACAGAGGCC-CCAGTTACTACTGGTGTAAATTTGTAGTTTCAGCTGGAGTCATCT 85
QY 1478 AGCTACTTAAGTCTGAGAAATATGCGAAGCAGCAGGAGCAGTCCAGTCAGGCGGC 1537
Db 86 AGCTCACTTAAAGTCTGAGAAATATGCGAAGCAGCAGGAGCAGTCCAGTCCAGGCGGC 145
QY 1538 CTTAATCCAGAGTTAATGATAGCTTTCCAGGGGCTGTGTCATTAAATC-AGATGATCT 1596
Db 146 CTTAATCCAGAGTTAATGATAGCTTTCCAGGGGCTGTGTCATTAAATCAGAGATGATCT 205
QY 1597 AGCCAAAGTTTCAGAGGAGAGACATGACCAAGACAGACAGAGCTCGAGTGGGGCTAA 1656
Db 206 ANCCNAAAGTTTCAGAGGAGAGACATGACCAAGACAGACAGAGCTCNAGTGGGGCTAA 265
QY 1657 AAGAGAGGACTTTCAGCAAGAGGAG-CCGAGGATAATATGTTACGAGCTATTAATGTC 1715
Db 266 AAGAGAGGACTTTCAGCAAGAGGAGCCCGGAGATAATATGTTACGAGCTATTAATGNC 325
QY 1716 AGTTCTGTGACTTCCGATATCCAAAGCCATGGCCCTGATTAATTTAGTGGGGCCAC 1775
Db 326 AGCCCTGTGACTTCCGATATCCAAAGCCATGGCCCTGATTAATTTAGTGGGGCCAC 385
QY 1776 TTCTCGTGTATATCAACAGCTCCATAACATTCACAA 1812
Db 386 TTCNCGTNAATCAACAGCTNGCATTACATTCACAA 422

RESULT 10
US-09-918-995-10414
; Sequence 10414, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10414
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: (1)-(476)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-10414

Query Match 7.2%; Score 322.8; DB 11; Length 476;
Best Local Similarity 96.2%; Pred. No. 1.9e-82;
Matches 327; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4162 AGTGAGTAAATGGTCAATGAGAAAGTGAAGATGGTAAACATCACTTTTAAACC 4221
Db 24 AGNGCANTCNNGNCGAGAAAGTGAAGATGGTAAACATCACTTTTAAACC 83
QY 4222 TGTTAAGTCAAAACCATCTTGGCTAATATGTTAGTGGGAAATTAATCCATAGAGATATCA 4281
Db 84 TGTTAAGTCAAAACCATCTTGGTAAATATGTTAGTGGGAAATTAATCCATAGAGATATCA 143
QY 4282 CCAGCTAGAAATTAATATTTTAAAGAAAGACCAAAATGTTAGAAATTTGAAAGG 4341
Db 144 CCAGCTAGAAATTAATATTTTAAAGAAAGACCAAAATGTTAGAAATTTGAAAGG 203
QY 4342 GTTTACATATTTATATTAATGAGCAGTCTGGCTGGCCATGGACCATTTGTTCCAAA 4401
Db 204 GTTTACATATTTATATTAATGAGCAGTCTGGCTGGCCATTTGGACCATTTGTTCCAAA 263
QY 4402 CCCATAAATTTGCTCTAAATTTTATAATGATCATGAACCCCTAGGCAGAGGAGGAAT 4461
```

```
Db 264 CCATATAATTTGGCTAAATTTATATGATCATGAACCCCTAGGCAGAGGAGGAAT 323
QY 4462 TGAAGGTCCAGGCAATGAAGAAATGGCCCTCTCA 4501
Db 324 TGAAGGTCCAGGCAATGAAGAAATGGCCCTCTCA 363

RESULT 11
US-09-778-320-110
; Sequence 110, Application US/09778320
; Patent No. US20030034052A1
; GENERAL INFORMATION:
; APPLICANT: Billion, Davin C.
; APPLICANT: Day, Craig H.
; APPLICANT: Jiang, Yugu
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Wang, TongTong
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 21021.491CS
; CURRENT APPLICATION NUMBER: US/09/778,320
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(305)
; OTHER INFORMATION: n = A,T,C or G
US-09-778-320-110

Query Match 6.1%; Score 275.4; DB 9; Length 305;
Best Local Similarity 94.1%; Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATATGACTTCCAGAGTGAAGCTGATGGCTGGTCTGGAGTAAATATAAGCTCTC 3498
Db 3439 ACATATGACTTCCAGAGTGAAGCTGATGGCTGGTCTGGAGTAAATATAAGCTCTC 60
QY 3499 CTTCTCTGGGAATCCGCACTACTTGAGTCACTGGCTGGCTTACCAATCTTCCAAAA 3558
Db 61 CTTCTCTGGGAATCCGCACTACTTGAGTCACTGGCTGGCTTACCAATCTTCCAAAA 120
QY 3559 CTATGTCCTTATCCAGCTT-CAATCTGGCTCTCATTTTTCAGCTTTGGATCAGACA 3617
Db 121 CTATGTCCTTATCCAGCTTNNAACTGNCCTCTCATTTTTCAGCTTTGGATCAGACA 180
QY 3618 ATGACATTTCTTAGATTTGGGATCAAGCATTCAGACCTGGGCAACTCCAAACGGTG 3677
Db 181 ATGACATTTCTTAGATTTGGGATCAAGCATTCAGACCTGGGCAACTCCAAACGGTG 240
QY 3678 CTCCAGGAGAAACGAGAGCCACCAATGTTAAAAATGAAGTCCCTTGAATGAG 3737
Db 241 CTTCNAGGAGAAACGAGAGCCACCAATGTTAAAAATGAAGTCCCTTGAATGAG 300
QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 12
US-09-510-689-110
; Sequence 110, Application US/09910689
; Patent No. US20020081509A1
; GENERAL INFORMATION:
; APPLICANT: Billion, Davin C.
; APPLICANT: Day, Craig H.
```

```
APPLICANT: Jiang, Yugu
APPLICANT: Houghton, Raymond L.
APPLICANT: Mitcham, Jennifer
APPLICANT: Wang, Tongtong
APPLICANT: McNeill, Patricia D.
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.491C6
CURRENT APPLICATION NUMBER: US/09/910,689
CURRENT FILING DATE: 2001-07-20
NUMBER OF SEQ ID NOS: 307
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 110
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: 12, 13, 16, 110, 142, 143, 150, 161, 192, 198, 217, 223,
LOCATION: 244, 263, 274, 285, 287
OTHER INFORMATION: n = A,T,C or G
US-09-910-689-110

Query Match      6.1%  Score 275.4; DB 9; Length 305;
Best Local Similarity 94.1%  Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGGTTCTGGAGTAAATATAGCTCTC 3498
Db 1 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGGTTCTGGAGTAAATATAGCTCTC 60

QY 3499 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCACAAA 3558
Db 61 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCACAAA 120

QY 3559 CTATGTGCCCTTATCCCACTT-CAATCTGCTCTCTCATTTTTCAGCTGTGGATCAGACA 3617
Db 121 CTATGTGCCCTTATCCCACTTNNAACTGNCCTCTCATTTTTCAGCTGTGGATCAGACA 180

QY 3618 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAACTGGTG 3677
Db 181 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAACTGGTG 240

QY 3678 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAGTCCCTTGAATGTAG 3737
Db 241 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAGTCCCTTGAATGTAG 300

QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 13
US-10-010-742-110
; Sequence 110, Application US/10010742
; Publication No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Day, Craig H.
; APPLICANT: Jiang, Yugu
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Wang, Tongtong
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Bennington, Angela Ann
; APPLICANT: Zehentner, Barbara
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.491C7
```

```
CURRENT APPLICATION NUMBER: US/10/010,742
CURRENT FILING DATE: 2001-11-30
NUMBER OF SEQ ID NOS: 307
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 110
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: 12, 13, 16, 110, 142, 143, 150, 162, 192, 198, 217, 223,
LOCATION: 244, 263, 274, 285, 287
OTHER INFORMATION: n = A,T,C or G
US-10-010-742-110

Query Match      6.1%  Score 275.4; DB 13; Length 305;
Best Local Similarity 94.1%  Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGGTTCTGGAGTAAATATAGCTCTC 3498
Db 1 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGGTTCTGGAGTAAATATAGCTCTC 60

QY 3499 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCACAAA 3558
Db 61 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCACAAA 120

QY 3559 CTATGTGCCCTTATCCCACTT-CAATCTGCTCTCTCATTTTTCAGCTGTGGATCAGACA 3617
Db 121 CTATGTGCCCTTATCCCACTTNNAACTGNCCTCTCATTTTTCAGCTGTGGATCAGACA 180

QY 3618 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAACTGGTG 3677
Db 181 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAACTGGTG 240

QY 3678 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAGTCCCTTGAATGTAG 3737
Db 241 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAGTCCCTTGAATGTAG 300

QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 14
US-10-198-846-7424
; Sequence 7424, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinhart, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; FILE REFERENCE: MFI-049
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7424
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2, 3, 4, 5, 6, 13, 434, 435, 436, 437, 438, 439, 440, 441,
; LOCATION: 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453,
; LOCATION: 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465,
; LOCATION: 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476
```

Search completed: October 29, 2003, 16:03:08
Job time : 123 secs